United States Patent [19]

Studier et al.

[11] **Patent Number: 4,952,496**

[45] **Date of Patent: Aug. 28, 1990**

[54] **CLONING AND EXPRESSION OF THE GENE FOR BACTERIOPHAGE T7 RNA POLYMERASE**

[75] Inventors: **F. William Studier**, Stony Brook, N.Y.; **Parichehre Davanloo**, Basel, Switzerland; **Alan H. Rosenberg**, Setauket, N.Y.; **Barbara A. Moffatt**, East Lansing, Mich.; **John J. Dunn**, Bellport, N.Y.

[73] Assignee: **Associated Universities, Inc.**, Washington, D.C.

[21] Appl. No.: **2,408**

[22] Filed: **Dec. 29, 1986**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 595,016, Mar. 30, 1984, abandoned.

[51] **Int. Cl.**[5] **C12P 19/34**; C12N 15/00; C12N 7/00; C12N 9/12

[52] **U.S. Cl.** **435/91**; 435/172.3; 435/194; 435/235; 435/252.3; 435/320; 536/27; 935/14; 935/29; 935/31; 935/43

[58] **Field of Search** 435/68, 70, 91, 653, 435/235, 240.1, 176.3, 317.1, 320; 536/27; 935/6, 29, 32, 34, 38, 41, 50, 57, 60

[56] **References Cited**

PUBLICATIONS

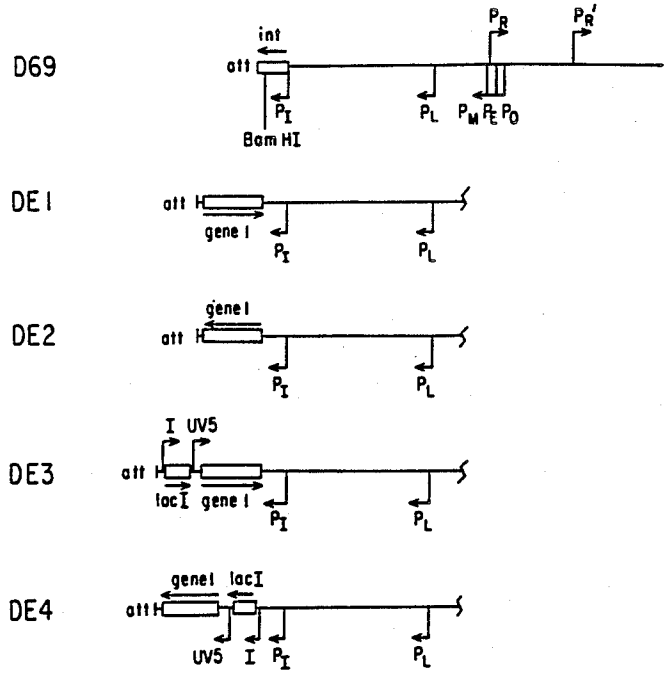
Green et al. (1983) Cell vol. 32, pp. 681–694.
Studier et al. (1983) Cold Spring Harbor Symp. Quant. Biol. XLVII: 999–1007.
Stahl et al. (1981) J. Mol. Biol., vol. 148, pp. 481–485.
Calos (1978) Nature vol. 274, pp. 762–765.
Maniatis et al. (1982) Cold Spring Harbor Labs., pp. 413–417 in *Molecular Cloning:* Aldercolor Manual.
McAllister et al. (1981) J. Mol. Biol. 153, pp. 527–544.
Studier et al. (1981) J. Mol. Biol. 153, pp. 503–525.
Morris, et al., Gene, vol. 41, pp. 193–200 (1986).
Simon, et al., J. Mol. Biol., vol. 79, pp. 249–265 (1973).
Hyman, et al., Virology, vol. 57, pp. 189–206 (1974).
Studier, et al., J. Virology, vol. 19, No. 1, pp. 136–145 (1976).
Hyman, et al., J. Mol. Biol., vol. 71, pp. 573–582 (1972).
Studier, et al., J. Mol. Biol., vol. 135, pp. 917–937 (1979).
Rosa, Cell, vol. 16, pp. 815–825 (1979).
Chamerlin, et al., Methods in Enzymbolgy, vol. 101, pp. 540–568 (1983).
Bailey, et al., Proc. Natl. Acad. Sci. U.S.A., vol. 80, pp. 2814–2818 (1983).
Studier, J. Mol. Biol., vol. 79, pp. 237–248 (1973).

Primary Examiner—Thomas Mays
Attorney, Agent, or Firm—Margaret C. Bogosian

[57] **ABSTRACT**

This application describes a means to clone a functional gene for bacteriophage T7 RNA polymerase. Active T7 RNA polymerase is produced from the cloned gene, and a plasmid has been constructed that can produce the active enzyme in large amounts. T7 RNA polymerase transcribes DNA very efficiently and is highly selective for a relatively long promoter sequence. This enzyme is useful for synthesizing large amounts of RNA in vivo or in vitro, and is capable of producing a single RNA selectively from a complex mixture of DNAs. The procedure used to obtain a clone of the T7 RNA polymerase gene can be applied to other T7-like phages to obtain clones that produce RNA polymerases having different promoter specificities, different bacterial hosts, or other desirable properties. T7 RNA polymerase is also used in a system for selective, high-level synthesis of RNAs and proteins in suitable host cells.

22 Claims, 6 Drawing Sheets



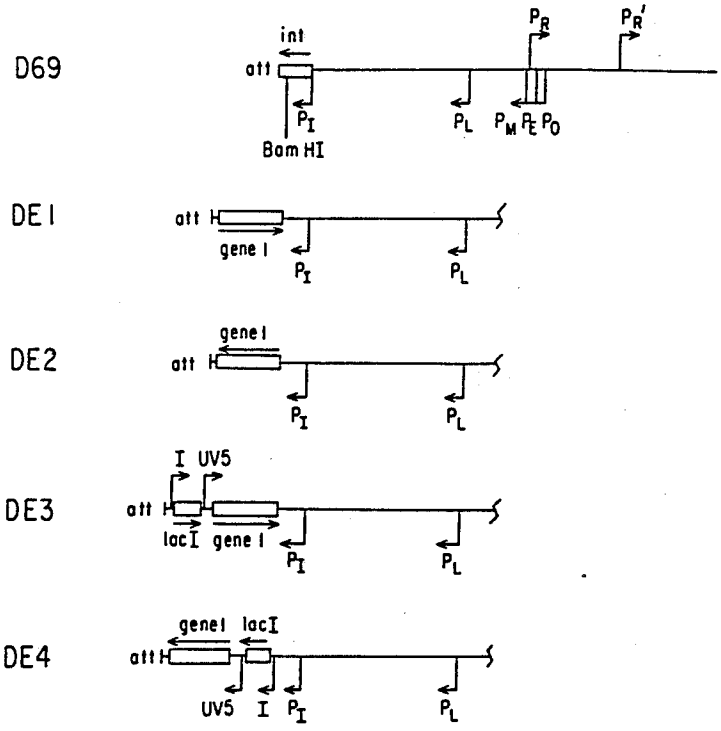

D69
int
att
P_I
Bam HI
P_R
P_R'
P_L
P_M P_E P_O
DE1
att
gene I
P_I
P_L
DE2
gene I
att
P_I
P_L
DE3
I UV5
att
lacI
gene I
P_I
P_L
DE4
gene I
lacI
att
UV5
I
P_I
P_L

Fig. 1

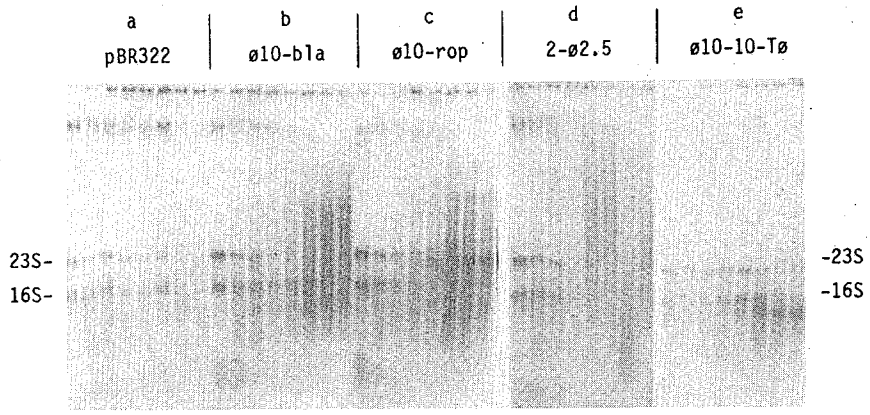

a
pBR322
b
ø10-bla
c
ø10-rop
d
2-ø2.5
e
ø10-10-Tø
23S-
16S-
-23S
-16S

Fig. 2

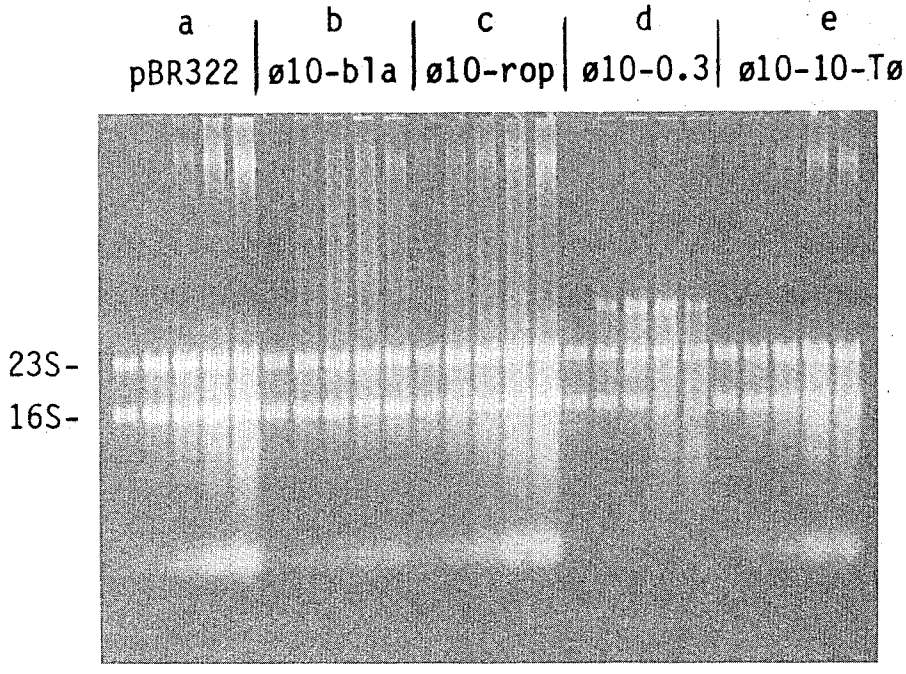

a
pBR322
b
ø10-bla
c
ø10-rop
d
ø10-0.3
e
ø10-10-Tø
23S-
16S-

Fig.3

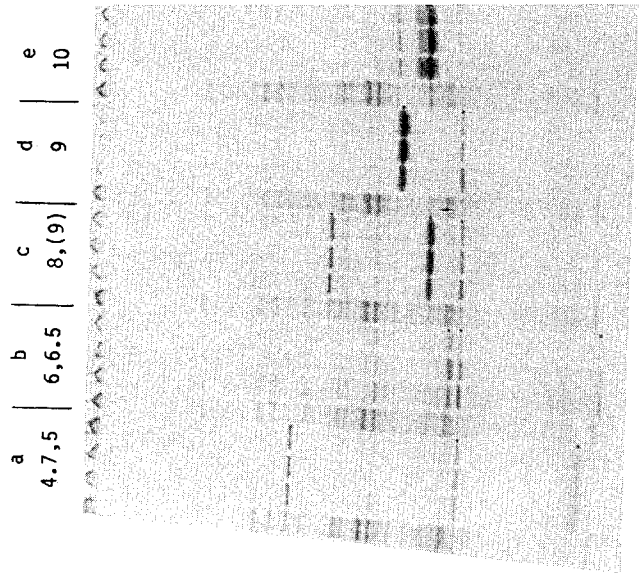


Fig.5

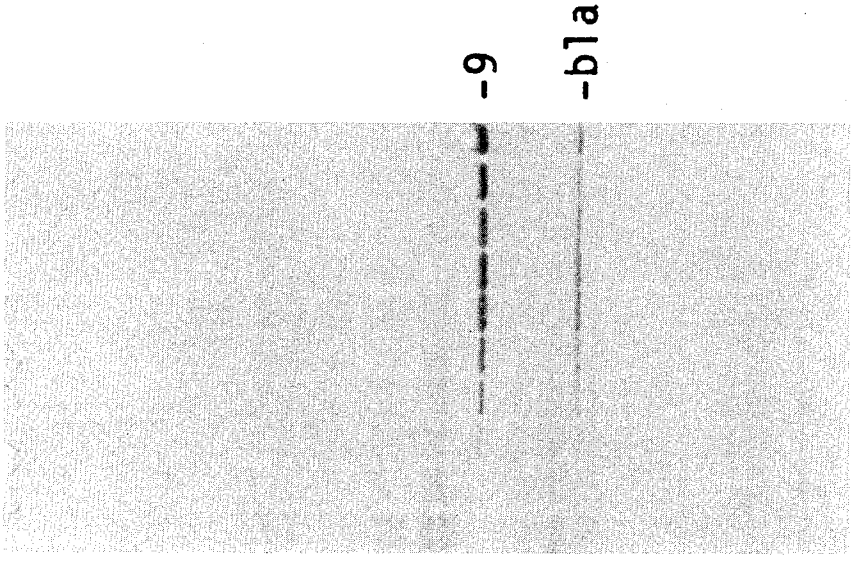


Fig.4

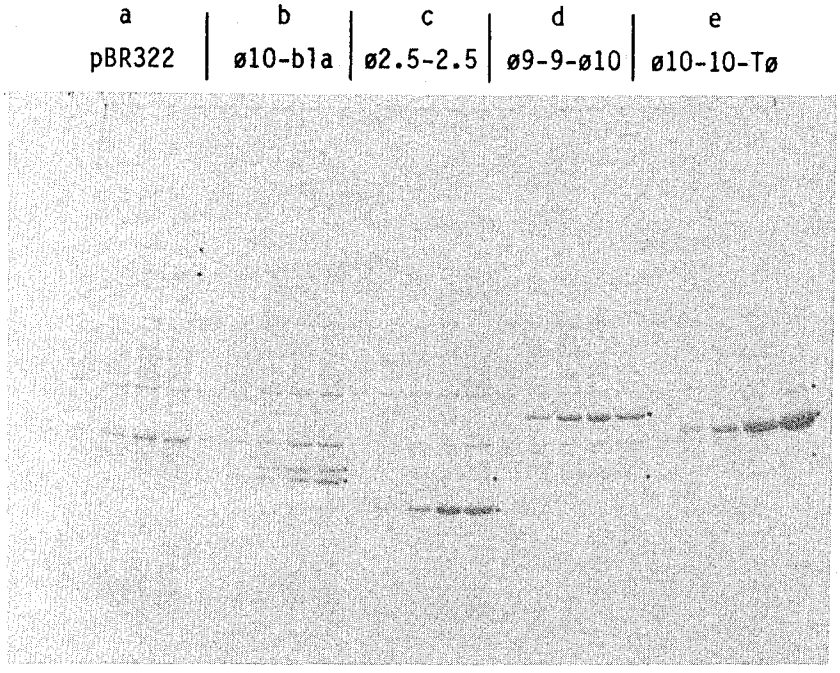


Fig.6

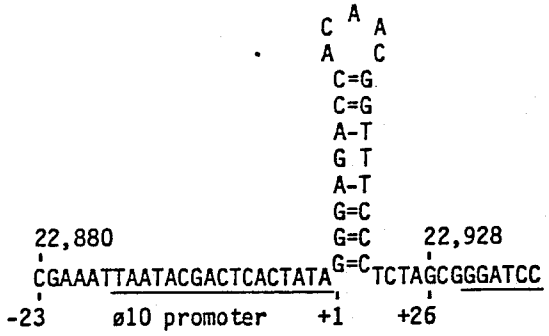

C A A
A C
C=G
C=G
A-T
G T
A-T
G=C
G=C
G=C
22,880
22,928
CGAAATTAATACGACTCACTATA TCTAGCGGGATCC
-23
ø10 promoter
+1
+26

Fig. 7

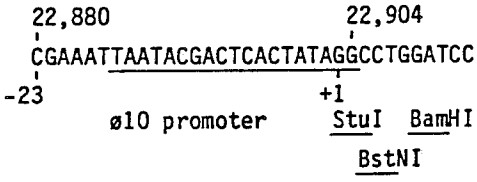

22,880
22,904
CGAAATTAATACGACTCACTATAGGCCTGGATCC
-23
+1
ø10 promoter
StuI
BamHI
BstNI

Fig. 8

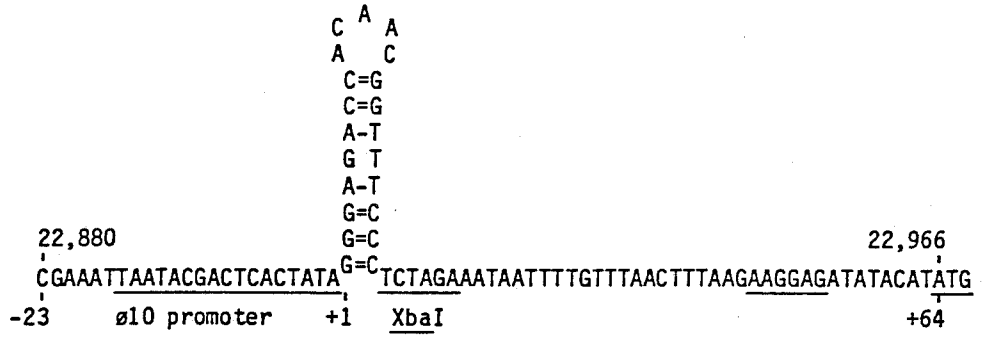

C A A
A C
C=G
C=G
A-T
G T
A-T
G=C
G=C
G=C
22,880
22,966
CGAAATTAATACGACTCACTATA TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATG
-23
ø10 promoter
+1
XbaI
+64

Fig. 9

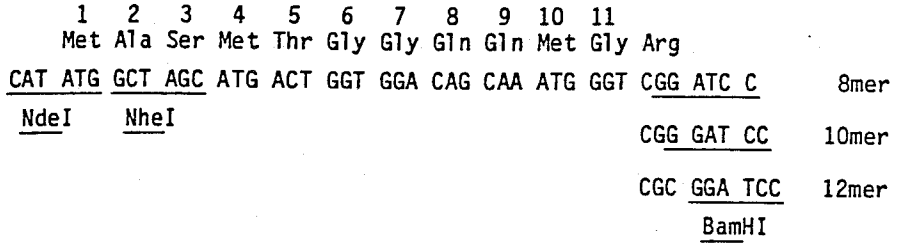

1 2 3 4 5 6 7 8 9 10 11
Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg
CAT ATG GCT AGC ATG ACT GGT GGA CAG CAA ATG GGT CGG ATC C 8mer
NdeI
NheI
CGG GAT CC 10mer
CGC GGA TCC 12mer
BamHI

Fig. 10

CLONING AND EXPRESSION OF THE GENE FOR BACTERIOPHAGE T7 RNA POLYMERASE

The U.S. Government has rights in this invention pursuant to Contract Number DE-AC02-76CH00016, between the U.S. Department of Energy and Associated Universities Inc.

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of co-pending application Ser. No. 595,016 filed Mar. 30, 1984, and now abandoned.

BACKGROUND OF THE INVENTION

Bacteriophage T7 is a virulent bacteriophage that infects *Escherichia coli*. It belongs to a class of bacteriophages that specify relatively simple RNA polymerases that selectively transcribe the DNA of their own bacteriophage but do not transcribe unrelated DNAs [Hausmann, *Current Topics in Microbiology and Immunology*, 75, 77–110 (1976); Korsten, et al., *J. Gen. Virol.*, 43, 57–73 (1979); Towle, et al., *J. Biol Chem.*, 250, 1723–1733 (1975); Butler and Chamberlin, *J. Biol. Chem.*, 257, 5772–5778 (1982); Chamberlin, et al., *Nature*, 228, 227–231 (1970); Dunn, et al., *Nature New Biology*, 230, 94–96 (1971)]. The T7 bacteriophage has been the subject of extensive scientific inquiry, in part because of its simple yet highly specific RNA polymerase. The genetic organization of T7 and the pattern of gene expression during infection are well understood, and the entire nucleotide sequence of T7 DNA is known (Studier and Dunn, *Cold Spring Harbor Symp. Quant. Biol.*, 47, 999–1007 (1983); Dunn and Studier, *J. Mol. Biol.*, 166, 477–535 (1983); Moffatt, et al., *J. Mol. Biol.*, 173, 265–269 (1984)]. These papers review and provide further references to a considerable body of work on T7 RNA polymerase.

T7 RNA polymerase, the product of T7 gene 1, is a protein produced early in T7 infection; it is a single-chain enzyme with a molecular weight close to 100,000. It appears that the basis for the selectivity of the T7 RNA polymerase is the interaction of the RNA polymerase with a relatively large promoter sequence, a sequence large enough that it is unlikely to be found by chance in any unrelated DNA. In the case of T7, the highly conserved promoter sequence appears to consist of approximately 23 continuous base pairs, which includes the start site for the RNA chain. If exact specification of even as few as 15 of these base pairs were required for initiation of chains, chance occurrence of a functional promoter would be expected less than once in a billion nucleotides of DNA.

The stringent specificity of T7-like RNA polymerases for their own promoter sequences is used by these phages to direct all transcription and replication to their own DNAs during infection. After the phage RNA polymerase is made, other phage gene products inactivate the host RNA polymerase, leaving all transcription in the cell directed by the phage enzyme.

T7 RNA polymerase is very efficient at transcribing DNA from its own promoters, and elongates RNA chains about five times faster than does *E. coli* RNA polymerase [Golomb and Chamberlin, *J. Biol. Chem.*, 249, 2858–2863 (1974)]. Termination signals for T7 RNA polymerase do not seem to occur very frequently,

and termination is usually not very efficient [McAllister, et al., *J. Mol. Biol.*, 153, 527–544 (1981)].

Their selectivity, activity, and ability to produce complete transcripts make T7 RNA polymerase and the equivalent RNA polymerases from T7-like phages useful for a variety of purposes. T7 RNA polymerase and SP6 RNA polymerase have been purified from infected cells and have been used to produce RNAs for translation in vitro [Dunn and Studier, *J. Mol. Biol.*, 166, 477–535 (1983)], substrates for splicing [Green, et al., *Cell*, 32, 681–694 (1983)], and hybridization probes [Zinn, et al., *Cell*, 34, 865–879 (1983)]. T7 RNA polymerase made during T7 infection directs the expression of genes under control of T7 promoters in plasmids [Campbell, et al., *PNAS, USA*, 75:2276–2280 (1978); Studier and Rosenberg, *J. Mol. Biol*, 153:503–525 (1981); and McAllister, et al., *J. Mol. Biol.*, 153:527–544 (1981)], but these gene products do not accumulate to high levels because of competition from promoters in T7 DNA and because the T7 infection quickly kills the cell. It was anticipated [McAllister, et al., *J. Mol. Biol.*, 153:527–544 (1981)], and the present invention demonstrates, that T7 RNA polymerase would also be useful for directing high-level expression of selected genes in cells. The problem in designing a high-level expression system is how to deliver active T7 RNA polymerase to a cell that contains a T7 promoter.

In the past, phage RNA polymerases like T7 RNA polymerase could be obtained only by infection with the phage from which they derive. The yield of purified RNA polymerase from infected cells is not particularly good, because the enzyme is synthesized for only a few minutes during the infection and does not accumulate to high levels. Nor is infection by these phages an efficient way to direct the transcription of non-phage genes inside the cell, because there is competition from promoters in the phage DNA itself and because the cells lyse within a short time.

Production of active T7 RNA polymerase from the cloned gene is an obvious way to obtain large amounts of enzyme for purification, and to have a source of enzyme that could be introduced into a variety of cells without the disadvantages associated with infection by T7 itself. Presumably recognizing this, other workers attempted to clone the active gene from T7 DNA but were not successful. In one report, Stahl and Zinn [*J. Mol. Biol.*, 148:481–485 (1981)] obtained a clone of the entire gene except for the last nucleotide of the termination codon. However, loss of the termination codon causes additional amino acids to be added to the carboxy terminus, and the protein produced from the clone was inactive.

The present invention discloses a successful process for cloning and expressing the T7 RNA polymerase gene, a process that can also be applied to clone the RNA polymerase genes from other T7-like phages. The same method has subsequently been applied by Tabor and Richardson, *PNAS, USA*, 82:1074–1078 (1985) to obtain a different clone of the T7 RNA polymerase gene, and by Morris, et al., *Gene*, 41:193–200 (1986) to obtain a clone of the T3 RNA polymerase gene. Having a clone of the active gene enables the use of it for making large amounts of RNA polymerase for purification, and also enables the use of it to direct sustained, high-level expression of selected genes in the cells. The present invention discloses successful methods for implementing these uses of the cloned gene for T7 RNA polymerase in *E. coli*.

UTILITY STATEMENT

The ability of T7 RNA polymerase and equivalent RNA polymerases from T7-like phages to transcribe selectively any DNA that is linked to an appropriate promoter can serve as the basis for a very specific and efficient production of desired RNAs both in vitro and inside a cell. RNAs produced in vitro can be useful as hybridization probes (for specific medical diagnosis, for example), as mRNAs for in vitro protein synthesis, as substrates for analyzing processing reactions or RNA splicing, or for any purpose requiring a specific RNA. RNAs produced inside the cell can direct the production of specific proteins of interest (antigens for vaccines, hormones, enzymes, or other proteins of medical or commercial value, for example), can form complexes with specific mRNAs to inhibit their translation selectively, in vitro or in the cell, or can be useful for any purpose requiring a specific RNA. Potentially, the selectivity and efficiency of the phage RNA polymerase could make such production very efficient. Furthermore, the unique properties of these phage RNA polymerases may make it possible for them to direct efficient expression of genes that are expressed only inefficiently or not at all by other RNA polymerases. These phage polymerases have the further advantage that it is possible to selectively inhibit the host cell RNA polymerase so that all transcription in the cell will be due to the phage RNA polymerase.

This invention discloses the first means of cloning the active gene for T7 RNA polymerase, a process that can also be used to clone the RNA polymerases of other T7-like phages. It also discloses a process whereby large amounts of T7 RNA polymerase can be expressed and purified, and processes whereby T7 RNA polymerase from the cloned gene, and potentially the RNA polymerases of any of the T7-like phages, can direct the production of large amounts of RNAs and proteins in cells.

DEPOSIT

Three plasmids of this invention have been deposited in the American Type Culture Collection prior to the filling date of this application and in accordance with the permanency and accessibility requirements of the U.S. Patent and Trademark Office. Plasmid pAR1151 has ATCC No. 39561; pAR1173 has ATCC No. 39562; and pAR1219 has ATCC No. 39563.

DESCRIPTION OF THE FIGURES

FIG. 1: Locations and orientations of *E. coli* promoters and inserted genes in D69, DE1, DE2, DE3 and DE4. These phages are also described in Table 1. The DNA between the attachment site (att) and the right end of the mature phage DNA is represented (for a map of the entire lambda DNA molecule see phage 473 of Hendrix et al., Lambda II, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. 1983). The locations of known *E. coli* promoters and the BamHI cloning site within the int gene of D69, and the locations and orientations of the lacI promoter, the lacI gene, the lacUV5 promoter, and T7 gene 1 in the different derivatives of D69 are represented to scale.

FIG. 2: Time-course of RNA synthesis after infection of different plasmid-carrying cultures by CE6. Cultures of HMS174 carrying different plasmids were grown in modified B2 medium containing 20 μg ampicillin/ml and infected with approximately 7 infectious particles of CE6/cell, as described in example 7. Samples of culture were labeled for 5 min with $^{32}PO_4$ (25 μCi/ml), and the cells were collected by centrifugation and suspended in an equal volume of 10 mM-sodium phosphate (pH 7.0), 2 mM-Na_3EDTA, 1% (w/v) sodium dodecyl sulfate, 1% (v/v) β-mercaptoethanol, 10% (v/v) glycerol, 0.01% (w/v) bromophenol blue. Each sample was mixed with 1/9 vol. 37% (v/v) formaldehyde, placed in a boiling waterbath for 2 min, and then an amount equivalent to 5 μl of culture was subjected to electrophoresis in a 1% (w/v) agarose gel containing 50 mM-sodium phosphate (pH 7.0), 2 mM-Na_3EDTA, 0.1% sodium dodecyl sulfate, followed by autoradiography. From left to right in each set, lanes represent samples labeled beginning 5 min before, and 0, 5, 10, 15, 30, 45 and 60 min after infection. The plasmids were: (a) pBR322, (b) pAR951, (c) pAR1494, (d) pAR219 and (e) pAR436. Where the plasmid contains any T7 promoters, T7 genes or Tφ, the order of these is indicated above the lanes, Whenever T7 genes are present, transcription from the T7 promoter is directed counterclockwise, which also directs transcription of the bla gene of the plasmid. When no T7 gene is present, the plasmid mRNA that would be transcribed from φ10 is indicated, bla for counterclockwise transcription and rop for clockwise. The positions of 23 S and 16 S ribosomal RNAs are indicated.

FIG. 3: Accumulation of RNA produced by T7 RNA polymerase. Cultures of BL21(DE3) carrying different plasmids were grown in M9ZB medium containing 20 μg ampicillin/ml, and synthesis of T7 RNA polymerase was induced by adding 0.4 mM-IPTG. The titer of a sample of culture removed before IPTG was added showed that almost all cells in each culture contained plasmid. Cells were collected by centrifugation, resuspended in 1.5 vol. 50 mM-Tris-HCl (pH 6–8), 2 mM-Na_3EDTA, 1% sodium dodecyl sulfate, 1% β-mercaptoethanol, 10% glycerol, 0.025% bromophenol blue, heated for 2 min in a boiling waterbath, and an amount equivalent to 3.3 μl of culture was subjected to electrophoresis in a 1.4% agarose gel in 40 mM-Tris-acetate (pH 8.0), 2mM-Na_3EDTA. RNA was visualized by ethidium bromide fluorescence. From left to right in each set, lanes represent samples collected immediately before, and 0.5, 1, 2 and 3 h after IPTG was added to the culture. The plasmids were: (a) pBR322, (b) pAR951, (c) pAR1494, (d) pAR946 and (e) pAR436. Genes under control of T7 promoters in the plasmid are indicated above each set, as described in the legend to FIG. 2. Plasmid pAR946 (set (d)) also contains the R0.3 RNase-III cleavage site. The positions of 23 S and 16 S ribosomal RNAs are indicated.

FIG. 4: Time-course of protein synthesis after infection of HMS174/pAR441 by CE6. A growing culture of HMS174/pAR441 was infected with approximately 14 infectious particles of CE6/cell, samples were labeled for 5 min with [^{35}S]methionine, and the labeled proteins were analyzed by gel electrophorosis followed by autoradiography. More than 99% of the cells in the culture were infected within 6 min, as indicated by loss of ability to form colonies. From left to right in the Figure, lanes represent samples that were labeled beginning at infection, and 5, 10, 15, 20, 30, 45, 60, 90 and 120 in after infection. The positions of the gene 9 and bla proteins are indicated.

FIG. 5: Time-course of protein synthesis after infection of different plasmid-carrying cultures by CE6. Cultures of HMS174 carrying different plasmids were grown, infected with approximately 8 to 12 infectious

particles of CE6/cell (except 25/cell for pAR1012), labeled for 5 min with [^{35}S]methionine, and the labeled proteins analyzed by gel electrophorosis followed by autoradiography. In each culture, more than 96% of the cells were infected within 6 min, as indicated by loss of ability to form colonies. From left to right in each set, lanes represent samples labeled immediately before, and 1, 2 and 3 h after infection. The plasmids were: (a) pAR1012, (b) pAR525, (c) pAR213, (d) pAR441 and (e) pAR436. The T7 proteins expressed under direction of T7 promoters in the plasmid are indicated above each set of lanes and marked by dots in the Figure; both the gene 10A and gene 10B proteins are marked. β-Lactamase is also expressed from T7 promoters in all of these plasmids, and its position is also marked in each set. The sizes of these proteins are given in Table 3, except for the fragment of gene 9 protein in set (c), which is predicted to contain 243 amino acid residues.

FIG. 6: Accumulation of protein directed by T7 RNA polymerase. Cultures of BL21(DE3) carrying the indicated plasmids were grown in ZY medium containing 200 μg ampicillin/ml, and synthesis of T7 RNA polymerase was induced by adding 0.4 mM-IPTG. The titer of samples of culture removed before IPTG was added showed that almost all of the cells had plasmid in each culture except that containing pAR511 (set (c)), where only about 87% of the cells had plasmid. Cells were collected by centrifugation, an amount equivalent to 10 μl of culture was subjected to gel electrophoresis, and the proteins were visualized by staining with Coomassie brilliant blue. From left to right in each set, lanes represent samples collected immediately before, and 0.5, 1, 2 and 3 h after IPTG was added to the culture. The plasmids were: (a) pBR322, (b) pAR951, (c) pAR511, (d) pAR441 and (e) pAR436. Genes under control of T7 promoters in the plasmid are indicated above each set, as described in the legend to FIG. 2. The positions of β-galactosidase (higher) and T7 RNA polymerase (lower), whose synthesis is induced by IPTG, are marked by dots in set (a). The positions of the proteins expressed from the plasmids under control of T7 promoters are marked in the other sets, including β-lactamase in sets (b) to (e) and both the gene 10A and 10B proteins in set (e).

FIG. 7: Nucleotide sequence of the cloned φ10 promoter (−23 to +26), showing positions of the upstream conserved sequence of the φ10 promoter, the RNA start, a potential stem-and-loop structure in the RNA, and the downstream BamHI cloning site.

FIG. 8: Nucleotide sequence of the cloned φ10 promoter (−23 to +2), showing the conserved promoter sequence and the downstream linking sequence containing StuI, BstNI, and BamHI sites.

FIG. 9: Nucleotide sequence from the beginning of the cloned fragment of T7 DNA to the initiation codon of gene 10, showing positions of the φ10 promoter, the RNA start, a potential stem-and-loop structure in the RNA, a unique XbaI site, and the Shine-Dalgarno sequence.

FIG. 10: Nucleotide and amino acid sequence at the beginning of the gene 10 protein in the plasmids used for expressing coding sequences directly from the initiation codon or as fusions to the gene 10 protein.

DETAILED DESCRIPTION OF THE INVENTION

T7 RNA polymerase is active and efficient, is selective for its own promoters, and will make complete transcripts of almost any DNA that is linked to a T7 promoter. These properties make the enzyme very useful, but in combination they are also responsible for the failure of previous workers to clone the active T7 RNA polymerase gene. We recognized that a specific promoter for T7 RNA polymerase lies just past the end of the coding sequence, and we reasoned that it might be difficult or impossible to clone a DNA fragment that contained both an active polymerase gene and an active promoter. If such a fragment were cloned in a circular plasmid, for example, the polymerase would be able, by initiating at the promoter and transcribing around the plasmid, to direct the synthesis of its own mRNA. This would lead to an autocatalytic increase both in the level of T7 RNA polymerase and in the rate of transcription of the plasmid, which almost certainly would be lethal to the cell. Potentially, a single molecule of active T7 RNA polymerase would be sufficient to trigger this response, so such a construction would be stable only if there were absolutely no expression of the cloned gene by the host RNA polymerase, a situation that is difficult or impossible to achieve.

Cloning of genes in general does not have to contend with the problem of a gene product that can act on one or more sequences that might also be cloned along with the gene, and perhaps that is why previous workers were unable to clone the active T7 RNA polymerase gene. The process we devised to circumvent this problem for the RNA polymerases of T7-like phages and to obtain a clone of the active RNA polymerase gene consists of three steps, which are illustrated in the case of T7 RNA polymerase:

(1) Identify the coding sequence for the RNA polymerase in the phage DNA. This can be done by standard techniques, as has been demonstrated for T7 RNA polymerase, and might include isolating and identifying amber mutations in the polymerase gene by their inability to make late RNAs or proteins [Studier, *Science*, 176:367–376 (1972)]; identifying the location of the gene in the phage DNA by testing for ability of mutants to recombine with cloned DNA fragments [Studier and Rosenberg, *J. Mol. Biol.*, 153:503–525 (1981)]; and if necessary by determining the nucleotide sequence [Dunn and Studier, *J. Mol. Biol.*, 166:477–535 (1983); Stahl and Zinn, *J. Mol. Biol.*, 148:481–485 (1981)].

(2) Identify the locations of the specific promoters for the phage RNA polymerase. Again in the case of T7, this has been done by transcribing the phage DNA or specific fragments of it with the phage RNA polymerase in vitro or in vivo [for example, Carter, et al., *J. Virol.*, 37:636–642 (1981); McAllister, at al., *J. Mol. Biol.*, 153:527–544 (1981)]. T7-like promoters typically have a conserved nucleotide sequence (23 continuous base pairs in the case of T7), and identification of this sequence would enable promoters to be identified in or near the nucleotide sequence of the RNA polymerase gene.

(3) Isolate and clone a DNA fragment that contains the entire coding sequence for the RNA polymerase but no active promoter. The critical step in the entire process is to obtain a DNA fragment containing the coding sequence for an active RNA polymerase but no active promoters for the same RNA polymerase. It is also useful, but may not be in all cases necessary to remove any promoters for the RNA polymerase(s) of the host cell in which the clone will be propagated. The promoters can be removed or inactivated by standard techniques that are widely applied. A procedure we used is

given in Example 1, but other commonly used procedures are equally applicable. For example, commonly used techniques for in vitro mutagenesis could be employed to introduce unique cleavage sites for restriction endonucleases at appropriate positions within cloned fragments containing the ends of the gene, the promoters could be removed by cleavage, and the intact gene without any promoters then reassembled from its separate fragments. In the unexpected event that a promoter for the phage RNA polymerase is located within the coding sequence for the RNA polymerase itself, this promoter would have to be inactivated by the now standard techniques of in vitro mutagenesis to eliminate promoter function without inactivating the gene. Once the nucleotide sequence of an RNA polymerase gene and its promoters is known, it would even be possible to synthesize a gene that will specify the correct amino acid sequence (or one that would make an active enzyme) but will not contain any active promoters.

The above process was used to clone the active gene for T7 RNA polymerase under conditions where its expression would be minimal, in case T7 RNA polymerase should be lethal. The BamHI site of the plasmid pBR322 was used as a cloning site because genes that are lethal if expressed to any significant extent had already been cloned in this site [Studier and Rosenberg, *J. Mol. Biol.*, 153, 503–525 (1981)]. To prevent expression of the cloned gene by the host RNA polymerase, a weak promoter for *E. coli* RNA polymerase that lies just ahead of the gene in the phage DNA was removed.

The above process will also provide active clones of the genes for similar RNA polymerases from other T7-like phages. T7-like phages are widely distributed in nature and infect many different types of bacteria. The genetic organization of all T7-like phages that have been examined has been found to be essentially the same as that of T7. Therefore, it is possible to isolate a DNA fragment containing the coding sequence for the RNA polymerase gene but not external promoters, in a way similar to that described for T7. Examples of T7-like phages include, but are not limited to *Escherichia coli* phages T3, φI, φII, W31, H, Y, A1, 122, cro, C21, C22, and C23; *Pseudomonas putida* phage gh-1; *Salmonella typhimurium* phage SP6; *Serratia marcescens* phages IV; Citrobacter phage ViIII; and Klebsiella phage No. 11 [Hausmann, *Current Topics in Microbiology and Immunology*, 75, 77–109 (1976); Korsten, et al., *J.. Gen. Virol.*, 43, 57–73 (1975); Dunn, et al., *Nature New Biology*, 230, 94–96 (1971); Towle, et al., *J. Biol. Chem.*, 250, 1723–1733 (1975); Butler and Chamberlin, *J. Biol. Chem.*, 257, 5772–5778 (1982)]. The RNA polymerases of other T7-like phages have selectivities for their own promoters that are comparable to the selectivity of T7 RNA polymerase for its promoters, and several different, non-overlapping or partially overlapping promoter specificities are known. Clones of RNA polymerase genes from T7-like phages that are already known (or could be isolated from nature), together with their specific promoters, could be useful for directing the expression of specific genes in bacteria other than *E. coli*. Clones of RNA polymerases having non-overlapping specificities might also be useful for controlling two or more different sets of genes independently in a single cell. Certain T7-like RNA polymerases might also have properties such as temperature, pH, or ionic strength optima, kinetic properties, or stability that would make them particularly useful or desirable for particular applications.

By our preferred design, the clone of the gene for T7 RNA polymerase contains no promoter for *E. coli* RNA polymerase ahead of the gene. This makes it possible to control the expression of the gene by inserting appropriate promoters immediately ahead of it. To facilitate this process, we placed a unique cloning site ahead of the coding sequence so that any inserted promoter plus the coding sequence for T7 RNA polymerase can be removed as a single fragment for insertion into vectors suitable for different cell types (pAR1173, Example 3). In order to be able to make large amounts of T7 RNA polymerase for purification from *E. coli*, we inserted the inducible lacUV5 promoter for *E. coli* RNA polymerase at this site (pAR1219, Example 4). When this promoter is induced, T7 RNA polymerase accumulates to a level of perhaps 10–20% of the cell protein. This enzyme is soluble and active, and a simple purification procedure yields about 10–15 mg of pure protein from 100 ml of induced culture (Example 5 and 6). This is an ample and convenient supply of purified T7 RNA polymerase for most purposes. Of course, much larger amounts of T7 RNA polymerase could be obtained by starting with larger amounts of induced culture, or perhaps by placing a different promoter ahead of the cloned gene.

The cloned gene provides a source of active T7 RNA polymerase inside *E. coli* cells without the disadvantages inherent in T7 infection. Since T7 RNA polymerase is highly selective for its own promoters, which do not occur naturally in *E. coli*, the presence of a T7 promoter in the same cell should direct all transcription by T7 RNA polymerase in the cell to the DNA controlled by that promoter. Potentially, the entire resources of the cell could be directed to the production of selected RNAs and proteins.

In attempting to exploit the clone of gene 1 for this purpose, we initially attempted to place all of the elements needed for this expression system in a single plasmid. Starting with plasmid pAR1219, which contains gene 1 under control of the inducible lacUV5 promoter, we tried to add target genes under control of a T7 promoter. These efforts were unsuccessful, apparently because the basal levels of active T7 RNA polymerase in the uninduced state were such that the plasmids could not be maintained.

T7 RNA polymerase is so active and selective that we considered it might be difficult to find conditions where cloned gene 1 would be shut off to such a degree that the cell could tolerate diverse target genes, many of which might be toxic. To provide a completely general solution to this problem, we decided to remove gene 1 from the cell entirely, and to introduce it only at the time we wished it to be active. In this way, we separate the system into two parts:

(1) the source of T7 RNA polymerase from the cloned gene, and

(2) the target gene under control of a T7 promoter.

In the two-part configuration, target genes could be cloned in the host cell under control of a T7 promoter without enhancing their expression, since *E. coli* RNA polymerase (and presumably the RNA polymerases of almost any potential host cell) does not initiate transcription at T7 promoters. If target genes are cloned in sites where the host RNA polymerase makes little or no mRNA from them, even genes whose products are very toxic to the cell should be tolerated under control of a T7 promoter. The gene for T7 RNA polymerase could be placed in a virus, but under control of promoters for

the host cell or viral RNA polymerase (if any). Upon infection with the virus, active T7 RNA polymerase would be produced and would in turn transcribe the target DNA that is under control of the T7 promoter.

A specific configuration for expression in which the target gene is cloned in the silent orientation in the BamHI site of pBR322 and the T7 RNA polymerase is delivered by a derivative of phage lambda is described in the examples. In this configuration, it is possible to direct essentially all of the resources of the cell to the production of target genes and proteins, and to accumulate large amounts of selected RNAs and proteins. This configuration was used to work out the properties of the T7 expression system, but it is not intended that the process be limited to this specific configuration. We expect that this process for gene expression will be successful with the target genes cloned in many different types of vectors or in the chromosome of the cell; with T7 RNA polymerase delivered by other types of virus besides lambda; with clones of other T7-like RNA polymerases, alone or in combination, together with target genes controlled by their specific promoters; and in other cell types besides *E. coli*, both prokaryotic and eukaryotic.

A complementary two-part configuration is also possible, in which the target gene is placed under control of a T7 promoter in the virus and the gene for T7 RNA polymerase is maintained in the cell under control of an inducible or constitutive promoter for the host RNA polymerase. In this configuration, it is possible to adjust the level of active T7 RNA polymerase present in the cell before delivery of the target gene by the virus. So far, the levels of expression produced in this configuration have not been as high as in the first process.

Although the properties of the T7 expression system were defined in the first configuration described above, it is also useful to be able to have a system where all parts are resident in the same cell. We have achieved such a unitary process as well. In this configuration a single copy of the gene for T7 RNA polymerase is placed in the chromosome of the cell under control of an inducible promoter for the host cell RNA polymerase and the target gene is placed under control of a T7 promoter in a multi-copy plasmid vector. With only a single copy of the T7 RNA polymerase gene in the cell, the basal activity in the uninduced state is low enough that many but not all target genes can be tolerated in the same cell. When the system is stable enough, the amounts of RNA and protein that can be produced are equally as large as in the first configuration above.

A specific unitary process for expression in which the target gene is cloned in the silent orientation in the BamHI site of pBR322 and the T7 RNA polymerase is placed in the chromosome as a lysogen of a derivative of phage lambda, and where the gene for T7 RNA polymerase is under control of the inducible lacUV5 promoter, is described in the examples. In this configuration, as in the first, it is possible to direct essentially all of the resources of the cell to the production of target genes and proteins, and to accumulate large amounts of selected RNAs and proteins. While this configuration is successful, it is not intended that the process be limited to this specific configuration. We expect that this process for gene expression will be successful with the T7 RNA polymerase gene cloned in the chromosome of the cell under control of many different inducible promoters; with the target genes cloned in many different types of single or multi-copy vectors or in the chromosome; with clones of other T7-like RNA polymerases, alone or in combination, together with target genes controlled by their specific promoters; and in other cell types besides *E. coli*, both prokaryotic and eukaryotic. In particular, since not much T7 RNA polymerase seems to be needed to produce very high levels of expression of target genes, inducible promoters that are very well shut off may be useful in allowing more toxic target genes to be tolerated even if the promoter is not very strong upon induction.

Conditions for optimal expression of target genes in *E. coli* were established using T7 genes under control of T7 promoters (examples 9–11). These examples show clearly that T7 RNA polymerase is capable of focusing the resources of the cell on the production of target proteins under control of a T7 promoter in a plasmid. The extent of the response depends upon the amount of T7 RNA polymerase present, but relatively small amounts generate a large response. The rate of synthesis of the target protein can be much higher than the rate of synthesis of any host protein, and such high rates can continue for three hours or longer. In any case, the large accumulation of RNA makes it seem likely that T7 RNA polymerase transcribing from a T7 promoter in a multicopy plasmid is capable of producing enough mRNA to saturate the protein-synthesizing machinery of *E. coli*.

Plasmid vectors have been constructed (examples 14 and 15) that allow DNA fragments from virtually any source to be placed under control of the strong $\phi 10$ promoter for T7 RNA polymerase, or under control of this promoter plus the protein initiation signals of T7 gene 10, the gene for the major capsid protein of T7. Some of these vectors also contain the transcription termination signal Tϕ from T7 DNA. The control sequences have been cloned so that they can easily be removed from these vectors and inserted in other plasmids, viruses or chromosomes to create a wide variety of other vectors and configurations or T7 RNA polymerase-directed expression of cloned DNA fragments both in vivo and in vitro. And of course, equivalent sets of vectors can also be created from the analagous control elements from other T7-like phages for use with their specific RNA polymerases.

The following examples provide additional elucidation of the plasmids and processes of this invention. Descriptions of the invention can also be found in Davanloo, et al., *PNAS, USA* 81:2035–2039 (1984) and in Studier and Moffatt, *J. Mol. Biol.*, 189:113–130 (1986). These examples are not intended to unduly restrict the invention to the uses described therein. In these examples, the following materials and methods were employed throughout:

1. *E. coli* HMS174 [as described in Campbell, et al., *Proc. Natl. Acad. Sci. USA*, 75, 2276–2280 (1978)] or BL21, a Met$^+$ derivative of B834 [described by Wood, *J. Mol. Biol.*, 16:118–133 (1966)] were used as the host for plasmid strains.

2. Bacteriophage T7, suitable hosts, and techniques for growing and manipulating them were used as described in Studier, *J. Mol. Biol.*, 94, 283–295 (1975); Studier, *J. Mol. Biol.*, 79, 227–236 (1973); and Studier, *Virology*, 39, 562–574 (1969). The nucleotide sequences, locations, and designations of T7 genes and genetic elements are given in Dunn & Studier, *J. Mol. Biol.* 166, 477–535 (1983) and Moffatt, et al., *J. Mol. Biol.* 173, 265–269 (1984).

3. DNA fragments were cloned in the plasmid pBR322, the plasmid described in Bolivar, et al., *Gene*, 2, 95–113 (1977).

4. The preparation and cloning of DNA fragments was by standard techniques as described in *Methods in Enzymology*, Volume 68, (1979), R. Wu, ed., Academic Press, New York, and *Molecular Cloning: A Laboratory Manual* (1982) by Maniatis, et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y., and Studier and Rosenberg, *J. Mol. Biol.*, 153, 503–525 (1981).

5. The source of lacI-lacUV5 promoter DNA was the plasmid pMCI described in Calos, *Nature* (*London*), 274, 762–765 (1978). Transcription from the lacUV5 promoter was induced by adding 0.4 mM isopropyl-beta-D-thiogalactoside (IPTG) to growing cultures.

6. Restriction endonucleases and enzymes used in cloning DNA were obtained from New England Biolabs, Bethesda Research Laboratories, or Boehringer Mannheim.

7. Synthetic deoxynucleotide linkers containing a BamHI site or a BglII site were obtained from New England Biolabs.

8. Preparation of the T7 Deletion Mutants: A number of deletions lying to the left or right of gene 1, the gene for T7 RNA polymerase, have been described and characterized [Studier, et al., *J. Mol. Biol.*, 135, 917–937 (1979)]. These deletions apparently arose by genetic crossovers between short repeated sequences. The crossover for D159 is at the sequence A-A-T-G-C-T-G-A, located at nucleotides 975 and 3023 in the nucleotide sequence of T7 DNA, and the crossover for C74 is at the sequence G-T-G-G-C-C-T, located at nucleotides 1458 and 3128 [Dunn and Studier, *J. Mol. Biol.*, 166, 477–535 (1983) and Studier, et al., *J. Mol. Biol.*, 135, 917–937 (1979)]. The likely crossover sequences for LG4 (A-A-T-A-C-G-A-C-T-C-A-C-T-A at 5832 and 7879) and LG26 (G-G-T-A-A-G-A-A at 7165 and 8658) were deduced from restriction mapping and the locations of repeated sequences in this region of T7 DNA. Heteroduplexes between the DNAs of double-deletion mutants D159,LG4 and C74,LG26, when digested with the single-strand-specific nuclease S1, would be expected to produce a fragment extending from nucleotides 3128 to 5845, which would contain the entire coding sequence for T7 RNA polymerase. The double-deletion strains were constructed by conventional genetic crosses, and the presence of both deletions in each double-deletion strain was confirmed by restriction analysis of the DNA.

9. Growth media: The practice of this invention includes, but is not limited to the following complex growth media: ZB medium (10 g N-Z-amine A/1 and 5 g NaCl/1); ZY medium 10 g N-Z-amine A/1, 5 g Bacto yeast extract/1, and 5 g NaCl/1); or M9 medium (1 g NH_4Cl/1, 3 g KH_2PO_4/1, 6 g $Na_2HPO_4$4/1, 4 g glucose/1, and 1 ml 1 m-$MgSO_4$/1); B2 medium (M9 medium in which all but 0.16 mM of the phosphate is replaced by salts and bis-Tris buffer); M9ZB medium (combines M9 and ZB media); M9 maltose and B2 maltose are the equivalent media, in which glucose is replaced by maltose. N-Z-amine A is commercially available from Sheffield Products; Bacto Tryptone and yeast extract are commercially available from Difco. When growing plasmid-containing cells, ampicillin was added to the medium, usually at a concentration of 20 ug/ml, but as high as 200 ug/ml. Dilutions of bacteria or phage for titering were made in ZB medium, and samples were plated by mixing with 2.5 ml of melted top agar [0.7% (w/v) agar in ZB medium], and spreading on plates containing 20 ml of hardened bottom agar (1% agar in ZB medium).

10. Maniatis, et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1982) discloses general procedures for working with lambda; the specific procedures followed for the present invention include: phage stocks were grown in ZY medium. Stock lysates were grown by adding a single plaque and 50 μl of a fresh overnight culture of ED8739 to 35 ml of growth medium in a 125 ml flask, and shaking at 37° C. until lysis; larger volumes were grown by adding 10 μl of lysate and 1 ml of cells to 500 ml of medium in a 1 liter flask. Lysates typically contained a few times 1010 infective phage particles/ml.

Phage were purified by precipitation with polyethylene glycol, followed by rapid isopycnic banding in CsCl step gradients. All solutions used during purification, including CsCl solutions, contained 10 mM-Tris Hcl (pH 8.0), 10 mM-$MgSO_4$, and 100 μg gelatin/ml to keep the phage intact. The purified phage were stored in the CsCl solution, and dilutions were made in 0.1 M-NaCl, 50 mM-Tris Hcl (pH 8.0), 10 mM-$MgSO_4$, and 100 μg gelatin/ml.

11. Cloning vector D69, the cloning vector used in the production of target gene proteins, is a lambda derivative that has imm^{21} and a single BamHI cloning site within the int gene. This vector is described in Mizusawa, et al., Gene, 20:317–322 (1982). A mixture of 500 ng of a BamHI digest of D69 DNA and a 5-fold molar excess of a BamHI fragment that contained the gene for T7 RNA polymerase (purified from pAR1151 or pAR1219) was ligated with phage T4 DNA ligase and packaged in a lambda packaging system. About 90% of the resulting plaques contained inserts (of both orientations). Cloning into the BamHI site of D69 interrupts the int gene, whose produce is needed for integration into the chromosome. Therefore, to form lysogens of the D69 derivatives that contain gene 1, int function was provided from a lysogen of heterologous immunity. A drop of lysate was spotted onto a lawn of the helper lysogen, the center of the cleared spot was used to grow a culture, and individual colonies from the culture were tested for the appropriate immunity or for the presence of a functional cloned gene. The genetic composition of D69 and the derivatives used in this process are shown in Table 1 and FIG. 1.

12. Rates of synthesis and accumulations of RNAs and proteins were analyzed by standard techniques of agarose and polyacrylamide gradient gel electrophoresis, as described in Studier, *J. Mol. Biol.*, 79:237–248 (1973 and *Molecular Cloning: A Laboratory Manual* (1982) by Maniatis, et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

EXAMPLE 1

Isolation of Gene 1 Fragments Lacking T7 Promoters

Search of the nucleotide sequence revealed no restriction sites that would permit convenient isolation of a fragment containing the entire gene 1 coding sequence but not the *E. coli* promoter located ahead of the coding sequence or the T7 promoter immediately following it. Therefore, an appropriate fragment was isolated from heteroduplexes formed between the DNAs of different deletion mutants of T7. The procedure is similar to that

employed by Stahl and Zinn, *J. Mol. Biol.*, 148, 481–485 (1981), except that heteroduplexes from overlapping double-deletion mutants were prepared in order to provide an optimal substrate for the single-strand specific endonuclease S1.

Following the techniques described in Studier, *J. Mol. Biol.*, 41, 199–209 (1969), heteroduplexes between the DNA of the double-deletion strains D159,LG4 and C74,LG26 were prepared directly from concentrated stocks of phage particles that had been purified by isopycnic banding in CsCl solution containing small amounts of Tris and EDTA. The stocks in CsCl solution were diluted 50 fold in 30 mM NaOH and left for 10 min. at room temperature to release and denature the DNA. The final DNA concentration was 100 μg/ml. The DNA mixture was neutralized by adding 0.1 volume of 0.5M Tris-Cl, pH 6.8, and the DNA was renatured by incubating 10 min. at 65° C. The renatured DNA was precipitated with ethanol and redissolved in 5 mM Tris-Cl, pH 6.8. The solution was adjusted to one-third the original volume and to a composition of 0.3M NaCl, 4.5 mM $ZnSO_4$, 30 mM sodium acetate, pH 4.6, for treatment with S1 nuclease. After digestion sufficient to release the double-stranded fragments from the heteroduplexes, the DNA mixture was subjected to electrophoresis on a 1% agarose gel, where the gene 1 fragment migrated as a sharp band well resolved from the other digestion products.

EXAMPLE 2

Construction of Plasmid pAR1151

Because of the possibility that a functional promoter for T7 RNA polymerase might remain in the fragment produced by S1 nuclease digestion of the heteroduplex DNA, we also digested the gene 1 fragments very lightly with BAL-31 exonuclease, to remove a few nucleotides from the ends of the fragment. The fragments generated by S1 alone, or by S1 followed by BAL-31 treatment, were purified by gel electrophoresis and adapted with BamHI linkers (CCGGATCCGG) for insertion into the BamHI site of pBR322. The ligation mixture was used to transform *E. coli* HMS174 to ampicillin resistance, and the transformants were enriched for tetracycline-sensitive clones by treatment with cycloserine.

When individual transformants from several different ligation mixtures were tested by gel electrophoresis of the plasmid DNAs, very few inserts were found, and none contained a complete gene 1. To enrich for plasmids containing an intact gene, plasmid DNA prepared from a mixed population of transformants obtained after cycloserine enrichment was subjected to gel electrophoresis, and DNA was recovered from the region of the gel where plasmids containing complete gene 1 would be expected to migrate. This DNA was then used to transform *E. coli* HMS174, and 48 new tetracylcine-sensitive clones were isolated and analyzed for inserts. Four of these 48 plasmid DNAs appeared to carry an insert of the proper size to contain all of gene 1. Gel electrophoresis of the fragments produced by cutting these plasmid DNAs with HindII and with KpnI produced identical patterns from all four clones, which were essentially the patterns expected if the entire gene had been cloned in the silent orientation. The fragments cloned in this experiment had been treated with BAL-31.

One of the plasmids produced by the above procedure which appeared to carry an insert of the proper size to contain all of gene 1 was pAR1151. To determine exactly where the cloned fragment begins and ends in the nucleotide sequence of T7 DNA, the fragment was released from the plasmid by cutting with BamHI and the nucleotide sequence at each end was determined by the techniques of Maxam and Gilbert (*Methods in Enzymology*, 65, 499–560, Academic Press, New York, 1979). The results showed that, after accounting for the sequence of the BamHI linkers, the fragment that was cloned in pAR1151 begins at nucleotide 3146 of T7 DNA and ends at nucleotide 5840. Because the last nucleotide of the linker at the left end of the fragment happens to be the same as nucleotide 3145 of T7 DNA and the first nucleotide of the linker at the right end the same as nucleotide 5841, the actual limits of T7 DNA sequence are 3145–5841, that is, position 7.88–14.63 in T7 DNA. Therefore, the cloned fragment contains slightly more than the entire coding sequence for T7 RNA polymerase.

Compared with the minimum fragment expected from S1 treatment of the heteroduplex DNA, the cloned fragment of pAR1151 has lost 19 base pairs from the left end and five from the right end, presumably because of the BAL-31 treatment. The first six nucleotides of the natural mRNA are missing, but the fragment does contain 26 base pairs ahead of the AUG initiation codon, and therefore has the entire predicted ribosome-binding and initiation region. The last 46 nucleotides of the natural mRNA are also missing, but the fragment contains 19 base pairs beyond the end of the coding sequence. The cloned fragment contains only the first 11 of the 23 highly conserved base pairs of the promoter for T7 RNA polymerase just past the end of the coding sequence and has no promoter activity for T7 RNA polymerase.

The plasmid pAR1151 was tested to determine if it produced active T7 RNA polymerase. When a cloned fragment of T7 DNA can express a functional gene, T7 amber mutants defective in that gene are usually able to plate with high efficiency on a restrictive host that carries the cloned fragment. In the case of gene 1 amber mutants, active T7 RNA polymerase is absolutely required to produce any increase in plating efficiency above the reversion frequency. We expected that little if any active T7 RNA polymerase would be produced from pAR1151, because the cloned gene is in the silent orientation in the BamHI site of pBR322, where little gene 1 mRNA should be produced. However, we found that gene 1 amber mutants form plaques on HMS174/pAR1151 with the same efficiency as on a host that carries an amber suppressor, although the plaques are somewhat variable in size. Clearly, active T7 RNA polymerase can be made in these cells. The variability of plaque size suggests that the level of active enzyme in the cell may be rather low but high enough that plaques are eventually produced from virtually every infection.

EXAMPLE 3

Construction of Plasmid pAR1173

In order to make large amounts of T7 RNA polymerase from the cloned gene, it was necessary to increase the synthesis of the gene 1 mRNA. To facilitate construction of plasmids in which promoters could be placed ahead of gene 1, a derivative of pAR1151 was first made in which a BglII site was inserted into the BamHI site ahead of gene 1. There are no BglII sites in

gene 1 or pBR322, so the inserted BglII site provides a unique cloning site immediately ahead of gene 1. The construction was made by partial digestion of pAR1151 with BamHI, filling in the ends with DNA polymerase, attaching synthetic BglII linkers (CAGATCTG), digesting with BglII, and religating the exposed BglII sites. Because of the sequence of the linkers, a BamHI site should be regenerated to each side of the BglII site. The modified plasmid selected, pAR1173, does not contain a BamHI site between the BglII site and gene 1, presumably due to some imperfection in the cutting and ligation reactions. This is a convenient result because sequences that are inserted in the BglII site will remain with gene 1 as a single fragment upon digestion with BamHI, allowing promoter-gene 1 constructs made in pAR1173 to be moved easily to other vectors.

EXAMPLE 4

Construction of Plasmid pAR1219 and Lambda Derivatives That Carry the T7 RNA Polymerase Gene

One promoter that has been inserted into the BglII site ahead of gene 1 in pAR1173 is the inducible lacUV5 promoter, which remains sensitive to the lac repressor but is not longer subject to catabolite repression [Silverstone, et al., PNAS, 66:773–779 (1970)]. The DNA fragment inserted contains the lacI gene as well as the lac UV5 promoter, eliminating the need to rely on the host cell's lac repressor to keep the promoter substantially shut off in the multicopy plasmid. This fragment was obtained from plasmid pMCl [Calos, *Nature London,* 274:762–765 (1978)]. The 1724 base-pair fragment contains the lacI gene, the lac UV5 promoter region, and the beginning of the lacZ gene isolated from pMCl by partial digestion with HincII and cloned into the BamHI site of pBR322 using synthetic linkers. The BamHI fragment isolated from this clone was then ligated into the BglII site of pAR1173 ahead of gene 1, taking advantage of the identity of the 4-base extensions produced by BglII and BamHI. A resulting plasmid in which the lac UV5 promoter is directed toward gene 1 is pAR1219. When induced with IPTG, this plasmid produces large amounts of active T7 RNA polymerase.

The active gene for T7 RNA polymerase, alone (pAR1151) or under control of the lac UV5 promoter (pAR1219), was placed in the BamHI site of the lambda cloning vector D69 (described in Mizusawa, et al., cited above). The locations and orientations of *E. coli* promoters and inserted genes in D69, DE1, DE2, DE3, and DE4 are shown in FIG. 1, and the phages are described in Table 1. All the phages contain both a mutation that eliminates the BamHI site at nucleotide 5505 of lambda DNA, and a deletion of the DNA between the EcoRI sites at nucleotides 21,226 and 26,104. The cloning site in D69 is the BamHI site at nucleotide 27,972 in the int gene. The orientation in which mRNA for T7 RNA polymerase would be transcribed from the P_L and P_I promoters of the phage is designated E, for expressed (in Table 1); the opposite orientation is designated S, for silent. The immunity region between nucleotides 34,379 and 38,617 either has been replaced by the immunity region of phage 21 or is the lambda immunity region having the cI857 and indl mutations. The nin5 deletion (Δ) removes nucleotides 40,502, to 43,307. In the orientation represented by DE1, no *E. coli* promoters that would direct transcription of gene 1 mRNA are known to be present in D69 DNA; DE3 contains gene 1 in the same orientation but under control of the lacUV5 promoter. In the opposite orientation, represented by DE2, the P_L and P_I promoters of D69 can direct transcription of gene 1 mRNA; DE4 contains gene 1 in this orientation but is also under control of the lac UV5 promoter. All four of these strains grow well and produce high-titer lysates, and all have been obtained as lysogens of HMS174 and BL21. Apparently, the presence of the T7 RNA polymerase gene, whether expressed or not, has little effect on lambda growth or lysogeny in the usual hosts.

Additional derivatives carrying gene 1 were obtained by crossing DE2 (in which gene 1 is under control of P_L and P_I) with lambda cI857indlSam7. The cI857 mutation would make expression from the P_L promoter temperature-inducible, and the Sam7 mutation would prevent lysis of the infected or induced cells. Three recombinants selected for further use all appear to have retained the left arm of DE2 and to have lost the nin5 deletion; DE6 retains imm^{21} and has acquired the Sam7 mutation; CE2 has acquired imm^{λ} (cI857indl); and CE6 has acquired both imm^{λ} (cI857indl) and Sam7.

TABLE 1

Derivatives of the Lambda Cloning Vector D69 that Carry T7 Gene 1

Phage	Fragment cloned into int	Orientation	Immunity	nin5	Sam7
D69	None		21	Δ	+
DE1	T7 gene 1	S	21	Δ	+
DE2	T7 gene 1	E	21	Δ	+
DE3	lacUV5-gene 1	S	21	Δ	+
DE4	lacUV5-gene 1	E	21	Δ	+
DE6	T7 gene 1	E	21	+	am
CE2	T7 gene 1	E	λcI857	+	+
CE6	T7 gene 1	E	λcI857	+	am

D69 is described by Mizusawa & Ward (1982). All phages in the Table contain both a mutation that eliminates the BamHI site at nucleotide 5505 of lambda DNA and a deletion of the DNA between the EcoRi sites at nucleotides 21,226 and 26,104. The cloning site in D69 is the BamHI site at nucleotide 27,972 in the int gene.

The orientation in which mRNA for T7 RNA polymerase would be transcribed from the pL and pI promoters of the phage (see FIG. 1) is designated E (for expressed); the opposite orientation is designated S (for silent).

The immunity region between nucleotides 34,379 and 38,617 either has been replaced by the immunity region of phage 21 or is the lambda immunity region having the cI857 and indl mutations. The nin5 deletion (Δ) removes nucleotides 40,502 to 43,307.

EXAMPLE 5

Induction of T7 RNA Polymerase from pAR1219

When cultures of HMS174/pAR1219 or BL21/pAR1219 were induced with IPTG, T7 RNA polymerase was produced at a rapid rate and accumulated to levels such that it was the major protein of the cell. Enzyme activity, as assayed in crude extracts, increased along with accumulation of the protein. The rate of increase of turbidity of the culture was almost as high for the induced culture as for a parallel uninduced culture, and colonies could form on plates containing inducer. This suggests that T7 RNA polymerase itself is not very toxic to the cells.

EXAMPLE 6

Purification of T7 RNA Polymerase from BL21/pAR1219

T7 RNA polymerase was isolated from induced cultures of BL21/pAR1219 growing in a shaking flask at 37° C. in a relatively rich medium [tryptone broth plus M9, Studier, *Virology,* 39, 562–574 (1969)]. IPTG was added to a final concentration of 0.4 mM when the cells reached a few times 10^8 per ml and shaking at 37° C. was

continued another 4 hrs. These cells contain perhaps 10–20% of the total protein as T7 RNA polymerase, virtually all of it soluble and active.

T7 RNA polymerase can be purified from the induced cells by conventional techniques, for example, Chamberlin, et al., *Nature,* 228, 227–231 (1970), and typically about 10–15 mg of essentially pure protein can be obtained from 100 ml of culture (about 1 g of packed cells). BL21 was used as the host cell because we have found it to lack a protease found in HMS174 (and in many laboratory strains of *E. coli*) that can nick the T7 RNA polymerase protein during purification. However, essentially intact T7 RNA polymerase can be purified from protease-containing strains if the protease is removed at an early stage of the purification.

The purified enzyme is very active and specific in transcribing DNA that has a promoter for T7 RNA polymerase, as demonstrated by electrophoretic analysis of the DNA template and the RNAs produced from it. Almost all of the RNA produced when purified enzyme was used to transcribe a mixture of two DNA fragments, one of which contained a promoter for T7 RNA polymerase and the other of which did not, was of the size expected if it had started at the promoter and ended at the end of the fragment. This indicates that the transcription was specific and the enzyme was free of RNase activity. At a high enzyme concentration all of the precursors were incorporated into RNA in 15 minutes; at a lower enzyme concentration, incorporation continued for at least an hour. In both cases, the DNA fragments remained intact, indicating that the enzyme was free of DNase activity.

EXAMPLE 7

T7 RNA Polymerase Provided By Infection

Each of the D69 derivatives listed in Table 1, except DE1, produces enough T7 RNA polymerase during infection to generate high-level expression of target genes cloned in a plasmid under control of a T7 promoter. The following procedure, using CE6 as the phage, produces efficient infection and high-level expression of the target gene protein.

Cultures are grown at 37° C., usually in a shaking incubator or waterbath, and plasmid-containing cells are grown in the presence of ampicillin, usually at 20 μg/ml. In order to obtain efficient infection (as measured by loss of colony-forming units), cultures are grown in the presence of maltose (and no glucose) to induce the lambda receptor. When the multiplicity of infection is higher than about 20 infectious phage particles per cell, the overall rate of protein synthesis after infection is drastically reduced and little protein is produced from the target gene. Protein synthesis was completely abolished in five different cultures after infection by 56 phage particles per cell, and almost completely abolished in three of them after infection by 28 particles per cell. All of the cultures retained active protein syntehsis at multiplicities of 7 and 14, and achieved high rates of synthesis of proteins whose genes are transcribed by T7 RNA polymerase. Inhibition of protein synthesis at high multiplicities is complete within five minutes after adding the phage. Because of this effect, optimal expression of target genes is achieved at multiplicities of 5 to 10—high enough to infect almost every cell, but low enough that protein synthesis is not inhibited.

If cells are grown in M9 maltose, rates of protein synthesis can be measured by pulse-labeling with [^{35}S]methionine (FIG. 5), and if grown in B2 maltose, rates of RNA synthesis can be measured by pulse-labeling with $^{32}PO_4$ (FIG. 2). When the absorbance at 500 nm (A_{600}) of an actively growing culture reaches 0.3, glucose is added to give a concentration of 4 mg/ml, and the culture is grown for an additional 1 to 2 hours, during which the A_{600} reaches 0.6 to 1, and the cell concentration is typically 5×10^8/ml. At this point, $MgSO_4$ is added to a final concentration of 10 mM, and purified CE6 phage is added to a final concentration of 2×10^9/ml (0.001 volume of a stock having an A_{260} of 6). The multiplicity of infection is typically around 7, and the efficiency of infection is typically greater than 95%, as measured by loss of colony-forming ability. Addition of glucose and 10 mM $MgSO_4$ to the medium is not necessary, but seems to give slightly better production of protein from the target genes. Cells are usually harvested 3 hours after infection, enough time for substantial accumulation of target protein, but not enough time for uninfected cells to overgrow the culture.

EXAMPLE 8

T7 RNA Polymerase Provided by Induction of a Chromosomal Copy of the Gene

The lysogen BL21 (DE3) contains a single copy of the gene for T7 RNA polymerase in the chromosome, under the control of the inducible lacUV5 promoter. Some T7 RNA polymerase is produced from the prophage even in the absence of added inducer, so plasmids containing toxic target genes can be difficult or impossible to maintain in BL21(DE3). However, when the plasmid can be maintained, addition of isopropyl-β-D-thiogalactopyranoside (IPTG) induces the lacUV5 promoter to produce T7 RNA polymerase, which in turn initiates high-level expression of the target gene in the plasmid. The level of expression is usually comparable with that found in Example 7 (by infection).

Because toxicity of the target gene can lead to loss of the plasmid or the accumulation of non-functional mutants, cultures are sampled just before induction and titered for viable cells on plates without additives, or in the presence of 0.5 mg of ampicillin, 2.5 mol of IPTG, or both, added to the top agar. BL21(DE3) carrying a plasmid without a T7 promoter will form colonies on each of these plates, as will non-functional mutants that retain plasmid. In a typical culture useful for producing target proteins, almost all cells will form colonies on plates without additives or containing only amicillin. Less than 2% of the cells will form a colony on plates containing only IPTG, and less than 0.01% will form a colony on plates containing both ampicillin and IPTG.

Cultures have been grown in M9, M9ZB, ZY medium, or ZY medium supplemented with 0.4% (w/v) glucose. Induction is with 0.4 mM IPTG, when the culture reaches an A_{600} of 1 (corresponding to about 5×10^8 to 10^9 viable cells per ml). Cells are typically harvested 3 hours after induction, after substantial accumulation of target protein is achieved, but before the culture can be overgrown with unproductive cells or cells that have lost their plasmid.

EXAMPLE 9

Plasmids Used for Working Out the Properties of the T7 Expression System

Plasmids used to establish the utility of T7 RNA polymerase for directing the transcription and translation of cloned DNA in *E. coli* are listed in Table 2.

TABLE 2

Plasmids	Genetic elements inserted
pBR322	
pAR951	ϕ10
pAR1494	ϕ10 (clockwise)
pAR946	ϕ10-RO.3-0.3
pAR219	2-ϕ2.5
pAR511	ϕ2.5-2.5
pAR1012	R4.7-ϕ4.7-4.7-5
pAR525	6-6.3-ϕ6.5-R6.5-6.5
pAR213	8-ϕ9-(9) (243aa)
pAR441	ϕ9-9-ϕ10
pAR436	ϕ10-10-Tϕ

All inserts are in the BamHI site of pBR322 and all, except in pAR1494, are oriented so that transcription from T7 promoters proceeds counterclockwise in the conventional representation of pBR322. The order of T7 genetic elements in the inserted fragments is given, including T7 promoters (ϕ2.5, ϕ4.7, ϕ6.5, ϕ9 or ϕ10), intact T7 genes (0.3, 2, 2.5, 4.7, 5, 6, 6.3, 6.5, 8, 9 or 10), RNase III cleavage sites (R0.3, R4.7, R6.5), and the Tϕ transcription terminator. The sizes and functions of proteins specified by the plasmids are given in Table 3. Plasmid pAR213 carries all of gene 8 and a large fragment of gene 9, predicted to direct a T7-pBR322 fusion protein that is 243 amino acid residues (aa) long.

TABLE 3

Sizes and Functions of Proteins Specified by Plasmids

Gene	No. of amino acids	M_r	Function
A. For pBR322 proteins			
bla precursor	285	31,393	
bla processed	263	28,899	Beta-lactamase
rop	63	7226	Control of replication
B. For T7 proteins			
0.3	116	13,678	Anti-restriction
2	63	7043	Anit-E. coli RNA polymerase
2.5	231	25,562	DNA-binding
4.7	135	15,208	Unknown
5	704	79,692	DNA polymerase
6	347	39,995	Exonuclease
6.3	37	4088	Unknown
6.5	84	9474	Unknown
8	535	58,989	Head-tail junction
9	306	33,766	Head assembly
10A	344	36,414	Major head protein
10B	397	41,800	Minor head protein

It is assumed that the bla precursor does not retain the initial methionine residue but that the rop protein does. The gene 10B protein of T7 is produced by frameshifting during translation of the 10A mRNA. The relative mobilities of T7 proteins in gel electrophoresis in the presence of sodium dodecyl sulfate do not always correspond to their relative molecular weights.

EXAMPLE 10

Production of RNA in *E. coli*

T7 RNA polymerase can direct very high level synthesis of RNA, using T7 promoters in plasmids in *E. coli*. Rates of synthesis and levels of accumulation can be comparable to those for ribosomal RNAs (FIGS. 2 and 3). Transcription from the T7 promoter can also interfere with RNA synthesis by *E. coli* RNA polymerase (FIGS. 2*b* and 2*c*). If no transcription termination signals for T7 RNA polymerase are in the plasmid, transcription can continue around the entire plasmid to produce heterogeneous RNAs larger than plasmid length (FIGS. 2*b*, 2*c* and 2*d*, and FIGS. 3*b* and 3*c*). Where the RNA contains a single efficient RNase III cleavage site, the large heterogeneous RNAs are cut at this site and accumulate as a discrete band of the length expected for transcription once around the plasmid (FIG. 3*d*). Within an hour after induction, the amount of RNA accumulated in this band can approach that found in the ribosomal RNA bands (FIG. 3*d*). Where the plasmid DNA contains the transcription termination signal Tϕ, transcription terminates efficiently to produce RNA of the expected length (FIGS. 2*e* and 3*e*). The RNAs appear to be relatively stable, and it seems likely that enough RNA accumulates to saturate the translation apparatus of *E. coli*.

Once active T7 RNA polymerase is present in the cell, transcription by the host cell RNA polymerase can be selectively inhibited. In the case of *E. coli*, this can be done by adding inhibitors such as rifampicin, which inhibits *E. coli* RNA polymerase but not T7 RNA polymerase, or by producing an inhibitor such as the gene 2 protein of T7 (FIG. 3*d*), which binds to *E. coli* RNA polymerase and inactivates it. This can be useful in applications where it is desirable or necessary that the transcripts from the T7 promoter be the only ones being produced in the cell.

EXAMPLE 11

Time Course and Rate of Target Protein Synthesis

T7 gene 9 protein is efficiently synthesized in *E. coli*, and plasmid pAR441 was used to analyze optimum conditions for protein synthesis directed by the T7 expression system. The -lactamase mRNA is also transcribed by T7 RNA polymerase in this plasmid. Upon CE6 infection of HMS174/pAR441, synthesis of gene 9 protein and -lactamase begins to increase 10 to 15 minutes after infection, and by 15 to 20 minutes after infection these proteins are the most rapidly synthesized in the cell. Their rate of synthesis continues to increase until at least 30 minutes after infection and remains at a very high level for at least another 90 minutes. During this period, the rate of synthesis of host proteins gradually declines, and there is little evidence for synthesis of lambda proteins. Synthesis of T7 RNA polymerase can be detected, but remains at a relatively low level; apparently, large amounts of T7 RNA polymerase are not needed in order to direct most of the protein synthetic capacity of the cell to the production of target proteins. The rate of synthesis of gene 9 protein is considerably higher than that of -lactamase, even though there is every indication that the two mRNAs are produced in comparable amounts and are comparably stable.

EXAMPLE 12

Target Protein Synthesis under Different Conditions for Delivery of T7 RNA Polymerase

Experiments similar to those shown in Example 11 have examined stimulation of synthesis of gene 9 protein and B-lactamase after infection of HMS174/pAR441 by D69, DE1, DE3, and DE4 (see FIG. 1 for the location of gene 1 relative to *E. coli* promoters in these phages). As expected, no stimulation of target protein synthesis

is observed after infection by D69, which can not supply any T7 RNA polymerase; nor is any stimulation observed after infection by DE1, a result indicating that little if any transcription of gene 1 occurs during infection by DE1. Stimulation by DE2, on the other hand, is comparable to that by CE6, as is stimulation by DE3 or DE4 in the presence of IPTG. In the absence of inducer, DE3 stimulates some target protein synthesis, but not as much as in the presence of inducer; apparently, the infecting copies of the lacUV5 promoter are only partially repressed. Likewise, DE4 stimulates considerably less target protein synthesis in the absence of inducer than in its presence or than is stimulated by DE2 (which contains no lacUV5 promoter); apparently, repression of the lacUV5 promoter in DE4 blocks transcription initiated at P_L and/or P_I. Finally, when DE2 infects an immune host, target protein synthesis is delayed by a few minutes and may not reach quite as high a level as in a sensitive host, but very substantial production of target protein occurs none the less. In an immune host, P_L should be repressed; perhaps more DNAs enter the cell than can be repressed by the cI protein present at the time of infection, or perhaps transcription from the P_I promoter, which is known to function at a low level in an immune host, is sufficient to produce this response.

Transcription of the target plasmid by T7 RNA polymerase is so active that it interferes with expression of lambda proteins during infection. As a result, the lambda infection is aborted and the cells do not lyse. Therefore, use of inhibitors of *E. coli* RNA polymerase, or infecting an immune cell to prevent lambda gene expression, or use of lambda mutants that cannot lyse the cell, are generally not needed nor useful for enhancing or extending target gene expression.

Comparably high levels of expression are obtained whether T7 RNA polymerase is delivered by a phage carrier or is induced from a chromosomal copy of the gene. Making a lysogen of DE3 is a convenient way to place a single copy of gene 1 into the chromosome under control of the lacUV5 promoter. Such lysogens are very stable because gene 1 is inserted into the int gene, whose produce is needed for excising the prophage from the chromosome. The basal level of T7 RNA polymerase in the uninduced single-copy lysogen is considerably lower than when the cell contains the multi-copy plasmids pAR1151 or pAR1219, as measured by ability to plate 4107, a deletion mutant of T7 that completely lacks gene 1. This low basal activity allows a wide variety of target genes under control of a T7 promoter to be tolerated in the cell. Target genes that are too toxic to be established in a DE3 lysogen can still be expressed by infection with a gene 1-carrying lambda derivative.

EXAMPLE 13

Production of different target proteins in *E. coli*

A variety of different T7 proteins having different degrees of toxicity to the host cell have been expressed in the T7 expression system, where T7 RNA polymerase has been delivered by infection with a lambda derivative or produced by induction from a chromosomal copy of gene 1. Protein patterns illustrating rates of target protein synthesis at different times after synthesis of T7 RNA polymerase was initiated are shown in FIG. 5, and patterns illustrating the accumulation of target proteins are shown in FIG. 6. Typically, the target proteins are produced at substantial rates, and synthesis of host proteins is greatly reduced. This shift in protein synthesis reflects the shift in mRNA population in favor of transcripts produced by T7 RNA polymerase, and the relatively high translational efficiency of the T7 mRNAs. High rates of synthesis of the target proteins can continue for at least three hours (as shown in FIG. 5) and have been observed to continue for at least seven hours in some of the cases we have examined. Target proteins can accumulate to become a substantial fraction of the total cell protein, perhaps 50% or more, within a few hours (FIG. 6). The relative rates of synthesis and accumulation of individual T7 proteins expressed in the T7 expression system differ considerably, and generally seem to parallel the relative rates of synthesis during T7 infection. It seems likely that the T7 expression system produces saturating amounts of mRNA, and that the relative rates of protein synthesis reflect mainly the relative efficiencies of translation of the individual T7 mRNAs.

EXAMPLE 14

Vectors for Transcribing Target DNAs

T7 RNA polymerase is potentially capable of transcribing any DNA that is placed under control of a T7 promoter. Table 4 lists a set of plasmid vectors that contain a strong promoter for T7 RNA polymerase followed by one or more unique cloning sites. In pAR2529, the cloning site is followed by Tϕ, the transcription terminator. These vectors are all derived from pBR322, and DNA from many sources can be cloned in them so that it can be transcribed by T7 RNA polymerase in vitro or in vivo. The DNA fragments containing the T7 promoter or T7 promoter and terminator have been designed so that they can easily be removed from certain of the vectors and used to create new expression sites in other plasmids, viruses, or chromosomes. The practitioner of the art will recognize that a wide range of different expression vectors could be created using these or similar DNA fragments that contain a T7 promoter or T7 promoter and terminator. Detailed description of some of the more useful vectors follows.

TABLE 4

Vectors for transcribing cloned DNAs				
Plasmid	Upstream site	T7 DNA relative to ϕ10 start	Cloning sites	Downstream
pAR1959	BamHI	−23 to +26	BamHI	
pAR2019		−23 to +26	BamHI	
pAR2305	BglII	−23 to +26	BamHI	
pAR2529	BglII	−23 to +26	BamHI	Tϕ BglII
pAR2192		−23 to +26	BamHI -EcoRI	Δin pBR322
pAR2369	BamHI	−23 to +2	StuI-BamHI	
pAR2463	BglII	−23 to +2	StuI-BamHI	

Plasmid pAR1959 was constructed by inserting a TaqI-XbaI fragment of T7 DNA (nucleotides 22,880–22,928) containing the ϕ10 promoter for T7 RNA polymerase into the BamHI site of pBR322 by using the synthetic linker CGGGATCCCG. The fragment extends from nucleotides −23 to +26 relative to the start of the RNA and is oriented so that transcription from the ϕ10 promoter is directed counterclockwise, opposite to transcription from the tetracycline promoter. Counterclockwise transcription through this region by *E. coli* RNA polymerase is low enough to permit relatively toxic genes to be cloned under control

of the T7 promoter. T7 RNA polymerase initiates active and selective transcription at the ϕ10 promoter both in vivo and in vitro.

To increase the usefulness and versatility of plasmids containing this ϕ10 promoter fragment, the following derivatives were constructed:

pAR2019, in which the upstream BamHI site of pAR1959 has been removed by opening, filling in, and re-ligating. This leaves the downstream BamHI site as a unique cloning site.

pAR2305, in which the upstream BamHI site of pAR1959 has been converted to a BglII site by opening, filling in, adding the linker GAGATCTC, cutting with BglII, and re-ligating. Both the upstream BglII site and the downstream BamHI site are unique in this plasmid. New ϕ10 vectors can be created by moving the BglII-BamHI fragment to a unique BamHI site.

pAR2529, in which the transcription terminator for T7 RNA polymerase, Tϕ, has been added just downstream of the BamHI cloning site of pAR2305. A fragment of T7 DNA containing Tϕ (nucleotides 24,106–24,228, where transcription terminates at nucleotide 24,209) was joined to the BamHI cloning site of pAR2305 through the sequence [GGATCC]GG-Tϕ-CCGGATCGAGATCTCGATCC, where the final C is nucleotide 375 in the BamHI site of pBR322. The downstream linker contains a BglII site, so the entire ϕ10-BamHI-T fragment can be removed from this plasmid as a BglII fragment for transfer to other vectors.

pAR2192, in which the DNA between the BamHI and EcoRI sites of pAR2019 (originally from pBR322) has been deleted to leave the BamHI and EcoRI sites in the sequence GGATCCGTTAAC. This deletion removes the tetracycline promoter, so transcription by *E. coli* RNA polymerase across these cloning sites should be relatively low in both directions. If a fragment cloned into the BamHI site has no EcoRI sites, cutting the plasmid at the EcoRI site will cause the RNA made by purified T7 RNA polymerase to end only a few nucleotides past the end of the cloned fragment.

In all of the above plasmids, the RNA transcribed from a cloned DNA will begin with 26 nucleotides of T7 sequence and a CG from the linker sequence before the first G of the BamHI cloning site. The first 21 nucleotides of the RNA could potentially fold into a relatively stable 8 base-pair stem-and-loop structure (FIG. 7).

In order to be able to make RNA with only two nucleotides preceding the transcript of a cloned DNA, we took advantage of the fact that nucleotides -1 to $+2$ of the ϕ10 promoter sequence are AGG, which is half of a StuI site (AGG'CCT). The cloned sequence downstream of $+2$ in pAR1959 and pAR2305 was replaced by the remaining half of the StuI site followed immediately by a BamHI site in the sequence CCTGGATCC to create pAR2369 and pAR2463 (FIG. 8). The StuI site is unique in these plasmids, and can be used to place any DNA fragment at position $+3$ by blunt-end ligation.

In spite of the replacement of nucleotides $+3$ to $+6$ of the conserved sequence of the ϕ10 promoter, T7 RNA polymerase initiates RNA chains efficiently at this promoter in both plasmids. We expect that most or all DNA fragments inserted at the StuI site will be transcribed by T7 RNA polymerase to produce RNAs having only an additional GG at their 5' end. Cutting at the downstream BamHI site can also be used to end transcription only a few nucleotides past the end of the cloned fragment.

The sequence CCTGG found in the StuI-BamHI sequence of pAR2369 and pAR2463 is a cleavage site for BstNI and also a methylation site for the dcm methylase of *E. coli*. Methylation at this site prevents cleavage by StuI, so plasmid must be prepared from a dcm$^-$ strain if it is to be cut efficiently at the StuI site. BL21 is a B strain of *E. coli* and as such is dcm$^-$.

EXAMPLE 15

Vectors for Transcribing and Translating DNAs in *E. coli*

Different mRNAs are translated with widely different efficiencies in *E. coli*. The factors affecting efficiency of translation are not completely understood, but coding sequences usually start with AUG, and the mRNA usually contains upstream sequences (the Shine-Dalgarno sequences) thought to be important in binding the mRNA to the ribosome and initiating translation. The structure of the mRNA is also thought to be important.

The major capsid protein of T7, specified by gene 10, is made very efficiently during T7 infection, much more rapidly than any host protein. The mRNA initiated at the ϕ10 promoter contains 63 nucleotides ahead of the gene 10 initiation codon (FIG. 9), and this leader sequence seems likely to be responsible at least in part for the efficiency of translation of the gene 10 mRNA. The mRNA begins with 21 nucleotides that could form a relatively stable stem-and-loop structure, and continues with an apparently unstructured, AT-rich region that contains a good Shine-Dalgarno sequence ahead of the initiation codon. The sequence containing the initiation codon forms part of an NdeI cleavage site in the DNA (CATATG). Specific cleavage of the DNA at this site can be used to fuse coding sequences directly to the initiation codon of gene 10. An NdeI cleavage site is particularly useful for this purpose because the initiation codon for any protein can potentially be made to be part of an NdeI cleavage site in its DNA by changing only upstream, noncoding nucleotides, and without changing any coding sequences for the protein. Fusing coding sequences to the leader sequence in such a way could place the initiation codon for any protein at a site known to be efficiently translated in *E. coli*.

We have constructed vectors containing the 10 promoter and gene 10 protein initiation region to a point just past the 11th codon for the gene 10 protein. Besides the NdeI cleavage site at the initiation codon, the sequence contains an NheI cleavage site at the second and third codons. In addition, we have placed BamHI cleavage sites after the 11th codon, using linker oligonucleotides of three different lengths. Such a set of three vectors makes it relatively easy to fuse any coding sequence after the 11th codon of the gene 10 protein so that it will be in the correct reading frame to be translated from the gene 10 initiation site. In one set of vectors, the BamHI cloning sites are followed by Tϕ, the transcription terminator. The DNA fragments containing the control sequences have been designed so that they can easily be removed from certain of the vectors and used to create transcription-translation sites in other plasmids, viruses or chromosomes. The practitioner of the art will recognize that a wide range of different expression vectors could be created using these or similar DNA fragments that contain a T7 promoter and a translation initiation

site, with or without a transcription terminator. Detailed description of some of the more useful vectors follows.

TABLE 5

Vectors for Transcribing and Translating Cloned DNAs

Plasmid	Upstream site	Expression signals	Fusion cloning site	Open frame	Downstream
pAR2075	BamHI	ϕ10-s10	BamHI (8mer)	ATC	
pAR2078	BamHI	ϕ10-s10	BamHI (10mer)	GAT	
pAR2084	BamHI	ϕ10-s10	BamHI (12mer)	GGA	
pAR2120		ϕ10-s10	BamHI (8mer)	ATC	
pAR2093		ϕ10-s10	BamHI (10mer)	GAT	
pAR2098		ϕ10-s10	BamHI (12mer)	GGA	
pAR2156	BglII	ϕ10-s10	BamHI (8mer)	ATC	
pAR2106	BglII	ϕ10-s10	BamHI (10mer)	GAT	
pAR2113	BglII	ϕ10-s10	BamHI (12mer)	GGA	
pAR3038	BglII	ϕ10-s10	BamHI (8mer)	ATC	Tϕ EcoRV
pAR3039	BglII	ϕ10-s10	BamHI (10mer)	GAT	Tϕ EcoRV
pAR3040	BglII	ϕ10-s10	BamHI (12mer)	GGA	Tϕ EcoRV

The parental plasmid for this set of plasmids is pAR2067, which was derived from pBR322 by eliminating the NdeI site at nucleotide 2298 by opening it, filling in, and re-ligating the blunt ends.

A TaqI-RsaI fragment of T7 DNA (nucleotides 22,880–22,998) containing the ϕ10 promoter for T7 RNA polymerase, the translation initiation site for the gene 10 protein has been inserted into the BamHI site of pAR2067. (Gene 10 specifies the major capsid protein of T7.) The fragment extends from nucleotides −23 to +96 relative to the start of the RNA and is oriented in each plasmid so that transcription from the ϕ10 promoter is directed counterclockwise, opposite to transcription from the tetracycline promoter. Initially, the fragment was cloned with a BamHI linker CGGGATCCCG (10 mer) attached to the upstream end and one of three BamHI linkers, CGGATCCG (8 mer), CGGGATCCCG (10 mer), or CGCGGATCCGCG (12 mer), attached to the downstream end of the T7 DNA fragment. This produced a set of three plasmids such that inserting a DNA fragment having an open reading frame into the downstream BamHI site will produce an in-frame fusion to the first 11 amino acids of the gene 10 protein in one of the three plasmids. If insertion is by means of a BamHI-compatible sticky end (GATC), the open reading frame generated by the 8 mer linker will be ATC, that by the 10 mer GAT, and that by the 12 mer GGA.

To make this set of plasmids convenient to use, the upstream BamHI site was eliminated by opening, filling in, and re-ligating, or was converted to a BglII site by adding the linker GAGATCTC to the filled in site, cutting with Bg1II and ligating. This leaves the downstream BamHI site as a unique cloning site in these plasmids. New vectors can be created by moving the BglII-BamHI fragment into any unique BamHI site.

A further modification has been to add Tϕ, the transcription terminator for T7 RNA polymerase, just downstream of the BamHI cloning site. A fragment of T7 DNA containing Tϕ (nucleotides 24,106–24,228, where transcription terminates at nucleotide 24,209) was joined to the BamHI cloning site through the sequence [GGATCC]GG-To-CCGGATATCC, where the final C is nucleotide 375 in the BamHI site of pBR322. The downstream linking sequence contains an EcoRV site (GAT'ATC), so the entire [ϕ10-s10-Tϕ] fragment can be removed from this plasmid as a BglII-EcoRV fragment for transfer to other vectors. (These plasmids also contain one other EcoRV site, at nucleotide 187 of pBR322.)

Each of the plasmids in this set contains a unique NdeI site (CA'TATG) that includes the initiating ATG for the gene 10 protein. Any coding sequence that also begins at an NdeI site can be joined to the upstream translation signals for the gene 10 protein, and such a fusion will specify only amino acids of the desired protein. Initiation codons that are not naturally part of an NdeI site could potentially be converted to such by directed mutagenesis.

The second and third codons for the gene 10 protein specify an NheI site (G'CTAGC), which can also be used to link coding sequences to the beginning of the gene 10 protein. (These plasmids also contain one other NheI site, at nucleotide 229 of pBR322.)

EXAMPLE 16

Expression of Various Target DNAs in Vectors for Transcribing or Translating Cloned DNAs

The vectors described in examples 14 and 15 have been used for transcribing and translating many different DNAs in vitro and in *E. coli* cells. Full length poliovirus cDNA was cloned into the EcoRI site of vector plasmid pAR2192 in both orientations and transcribed by purified T7 RNA polymerase to produce full length infectious poliovirus RNA and full length complementary RNA; and the full length cDNA was cloned in the StuI site downstream of the T7 promoter of pAR2369 and transcribed to produce full length, infectious poliovirus RNA having only two additional nucleotides at the 5′ end and no more than seven nucleotides past the poly(A) tract at the 3′ end, as described in van der Werf, et al., *PNAS*, USA 83, 2330–2334 (1986). DNA specifying the 23S and 5S rRNAs of the rrnB ribosomal operon of *E. coli* was cloned between the XbaI and BamHI sites of plasmid pAR2156 in the correct orientation for synthesis of the rRNAs from the T7 promoter, and DNA specifying the entire rrnB operon except for the natural promoters was cloned in the BamHI site of plasmid pAR2192 in both orientations; complete transcripts were made of each of these cloned DNAs by purified T7 RNA polymerase, and the ribosomal RNAs were produced in vivo in BL21 (DE3) by T7 RNA polymerase induced by IPTG from chromosomal copy of the T7 RNA polymerase gene under control of the lacUV5 promoter, as described in Steen, et al., *EMBO J.*, 5, 1099–1103 (1986). A fragment of poliovirus cDNA

coding for a protease involved in processing the poliovirus polyprotein was cloned between the BamHI and HindIII sites of plasmid pAR2106 to form an in-frame fusion after the 11th codon of the T7 gene 10 protein; active protease was produced in HMS174 when T7 RNA polymerase was delivered to the cell by infection with CE6, as described in Toyoda, et al., Cell, 45, 761–770 (1986). These are only a few examples of the many uses already made of the vectors described in examples 14 and 15 to direct transcription of DNA fragments by purified T7 RNA polymerase or to direct in vivo expression of DNA fragments by T7 RNA polymerase in BL21(DE3) (after induction) or in HMS174 (after infection by CE6).

What is claimed is:

1. A process for obtaining clones of a DNA sequence encoding an enzymatically active RNA polymerase of any T7-like bacteriophage, comprising the steps of:

obtaining from the DNA of said T7-like bacteriophage a fragment that contains the complete coding sequence of a gene for said bacteriophage RNA polymerase;

eliminating from the fragment, or inactivating, any promoters that may lie to the 5′ side of the 5′ terminus of said coding sequence and which could direct the synthesis by any host cell RNA polymerase of active mRNA from the cloned gene;

eliminating from the fragment, or inactivating, any promoters that lie to the 5′ side of the 5′ terminus, to the 3′ side of the 3′ terminus and between said termini of said coding sequence and which are recognizable by said phage RNA polymerase, but retaining within the DNA fragment a sequence encoding an enzymatically active phage RNA polymerase; and

cloning said DNA fragment containing a sequence encoding enzymatically active phage RNA polymerase, but not containing functional promoters recognizable by said phage RNA polymerase, in a plasmid or other vector that contains no promoters recognizable by said phage RNA polymerase, and at a site in the vector where functional mRNA for said phage RNA polymerase would be only minimally transcribed by any host cell RNA polymerase.

2. The process of claim 1 applied to bacteriophage T7 to obtain clones of a DNA sequence encoding an enzymatically active T7 RNA polymerase.

3. Plasmid pAR1151, ATCC 39561, containing a DNA sequence encoding enzymatically active T7 RNA polymerase, said plasmid comprising base pairs 3145 to 5841 of T7 DNA inserted into the BamHI site of pBR322 by use of synthetic linkers.

4. A BamHI-ended DNA fragment containing a sequence encoding enzymatically active T7 RNA polymerase which is obtainable from plasmid pAR1151, ATCC 39561.

5. Plasmid pAR1173, ATCC 39562, consisting of plasmid pAR1151 modified to contain a single BglII site to the 5′ side of the 5′ terminus of the DNA sequence encoding enzymatically active T7 RNA polymerase and within the BamHI-ended DNA fragment containing said coding sequence.

6. A BamHI-ended DNA fragment, containing a sequence encoding enzymatically active T7 RNA polymerase and a BglII site to the 5′ side of the 5′ terminus of said coding sequence and which is obtainable from plasmid pAR1173, ATCC 39562.

7. Plasmid pAR1219, ATCC 39563, consisting of a DNA fragment containing the lacI gene and the lacUV5 promoter of E.coli inserted into the BglII site of pAR1173 in the orientation such that transcription from the lacUV5 promoter produces functional mRNA encoding enzymatically active T7 RNA polymerase.

8. A BamHI-ended DNA fragment containing the lacI gene, the lacUV5 promoter, and a sequence encoding enzymatically active T7 RNA polymerase, and which is obtainable from plasmid pAR1219, ATCC 39563.

9. The process of claim 1 wherein the T7-like phages are selected from the group consisting of *Escherichia coli* phages T3, φI, φII, W31, H, Y, A1122, cro, C21, C22 and C23; *Pseudomonas putida* phage gh-1; *Salmonella typhimurium* phage Sp6; *Serratia marcescens* phage IV; Citrobacter phage ViIII; and Klebsiella phage number 11.

10. A process for producing enzymatically active RNA polymerase of a T7-like bacteriophage, comprising the steps of:

obtaining a DNA fragment containing a sequence encoding said enzymatically active bacteriophage RNA polymerase but not containing functional promoters recognizable by said bacteriophage RNA polymerase according to the process of claim 1;

placing said DNA fragment under control of a promoter that is effective for directing the synthesis of mRNA that can be translated to produce said enzymatically active bacteriophage RNA polymerase in a suitable host cell;

providing said appropriately controlled coding sequence to a host cell suitable for the production of said enzymatically active bacteriophage RNA polymerase; and

incubating under conditions that allow production, accumulation, and isolation of said enzymatically active bacteriophage RNA polymerase.

11. The process of claim 10 wherein the T7-like bacteriophage is selected from the group consisting of *Escherichia coli* phages T7, T3, φI, φII, W31, H, Y, A1122, cro, C21, C22, and C23; *Pseudomonas putida* phage gh-1; *Salmonella typhimurium* phage SP6; *Serratia marcescens* phage IV; Citrobacter phage ViIII; and Klebsiella phage number 11.

12. The process of claim 10 wherein the T7-like phage is T7; the cloned coding sequence is under control of the lacUV5 promoter in plasmid pAR1219; the host cell is BL21 or other suitable host cell; and conditions for production of enzymatically active T7 RNA polymerase include incubating said plasmid-containing host cell under conditions suitable for cell growth and production of proteins, and inducing the production of T7 RNA polymerase by adding a suitable concentration of IPTG or other inducer of the lacUV5 promoter.

13. A process for creating vectors to supply enzymatically active DNA polymerase of a T7-like bacteriophage to cells, comprising the steps of:

obtaining a DNA fragment containing a sequence encoding said enzymatically active bacteriophage RNA polymerase but not containing functional promoters recognizable by said bacteriophage RNA polymerase, according to the process of claim 1;

placing said DNA fragment in a selected vector and under control of a promoter that can direct the synthesis of mRNA that can be translated to pro-

duce said enzymatically active bacteriophage RNA polymerase in a desired host cell, the resulting vector being suitable for delivery to said desired host cell.

14. The process of claim 13 wherein said DNA fragment is the BamHI-ended DNA fragment from plasmid pAR1151, and said DNA fragment is inserted into the BamHI site of the lambda cloning vector D69.

15. Bacteriophage DE2, produced by the process of claim 14, and having a sequence encoding enzymatically active T7 RNA polymerase in the orientation to be expressed from the PI or PL promoters.

16. A process for creating vectors in which a DNA sequence encoding enzymatically active T7 RNA polymerase can be transcribed from a promoter recognizable by a selected host cell or viral RNA polymerase, comprising the steps of:

inserting said promoter into the BglII site of plasmid pAR1173, said promoter being oriented so as to produce mRNA encoding enzymatically active T7 RNA polymerase;

isolating the BamHI-ended DNA fragment containing said promoter and said coding sequence for T7 RNA polymerase; and

inserting said BamHI-ended DNA fragment into a selected vector suitable for establishment in or delivery to the desired host cell.

17. The process of claim 16 wherein said promoter is an inducible promoter.

18. The process of claim 17 wherein the BamHI-ended DNA fragment from plasmid pAR1219, is inserted into the BamHI site of the lambda cloning vector D69.

19. Bacteriophage DE3, produced by the process of claim 18, wherein the BamHI-ended DNA fragment from plasmid pAR1219 is in the orientation opposite to that in which functional mRNA encoding T7 RNA polymerase would be expressed from the PI or PL promoters.

20. A process for inserting the DNA sequence encoding enzymatically active T7 RNA polymerase into a bacterial chromosome, comprising the steps of:

making a lysogen of bacteriophage DE2 or DE3 in any bacterial cell capable of forming a lysogen of DE2 or DE3, under conditions that provide a source of active int protein.

21. The lysogens BL21(DE2), BL21(DE3), HMS174(DE2) and HMS174(DE3) formed by the process of claim 20.

22. Bacteriophage CE6, obtained from a genetic cross between DE2 and lambda cI857indlSam7, wherein said bacteriophage CE6 comprises a DNA sequence encoding enzymatically active T7 RNA polymerase and the cI857indlSam7 mutations.

* * * * *

US005869320A

United States Patent [19]

Studier et al.

[11] **Patent Number: 5,869,320**

[45] **Date of Patent: Feb. 9, 1999**

[54] **CLONING AND EXPRESSION OF THE GENE FOR BACTERIOPHAGE T7 RNA POLYMERASE**

[75] Inventors: **F. William Studier**, Stony Brook, N.Y.; **Parichehre Davanloo**, Basel, Switzerland; **Alan H. Rosenberg**, Setauket, N.Y.; **Barbara A. Moffatt**, Waterloo, Canada; **John J. Dunn**, Bellport, N.Y.

[73] Assignee: **Brookhaven Science Associates LLC**, Upton, N.Y.

[21] Appl. No.: **784,201**

[22] Filed: **Jan. 15, 1997**

Related U.S. Application Data

[63] Continuation of Ser. No. 259,560, Jun. 14, 1994, Pat. No. 5,693,489, which is a continuation of Ser. No. 874,386, Apr. 27, 1992, abandoned, which is a continuation of Ser. No. 542,056, Jun. 22, 1990, abandoned, which is a continuation-in-part of Ser. No. 002,408, Dec. 29, 1986, Pat. No. 4,952,496, which is a continuation-in-part of Ser. No. 595,016, Mar. 30, 1984, abandoned.

[51] **Int. Cl.**[6] **C12N 1/21**; C12N 15/34; C12N 15/63; C12N 15/70

[52] **U.S. Cl.** **435/252.33**; 435/252.3; 435/320.1; 435/69.1; 530/350; 530/412; 536/23.72

[58] **Field of Search** 435/320.1, 252.3, 435/69.1, 252.33; 530/350, 412; 536/23.72

[56] **References Cited**

U.S. PATENT DOCUMENTS

5,026,645 6/1991 Kotani et al. 435/194

5,037,745 8/1991 McAllister et al. 435/91.3

OTHER PUBLICATIONS

Fischer et al., J. of Biol. Chem. 255(16):7956–7964 (1980).

Bulter et al., J. of Biol. Chem. 257(10):5772–5778 (1982).

Studier et al., Cold Spring Harbor Symp. Quan. Biol. 47:999–1007 (1983).

Dunn et al., J. Mol. Biol. 166:477–535 (1983).

Moffatt et al., J. Mol. Biol. 173:265–269 (1984).

Davanloo et al., Proc. Natl. Acad. Sci. USA 81:2035–2039 (1984).

Primary Examiner—Johnny F. Railey, II
Attorney, Agent, or Firm—Margaret C. Bogosian

[57] **ABSTRACT**

This application describes a means to clone a functional gene for bacteriophage T7 RNA polymerase. Active T7 RNA polymerase is produced from the cloned gene, and a plasmid has been constructed that can produce the active enzyme in large amounts. T7 RNA polymerase transcribes DNA very efficiently and is highly selective for a relatively long promoter sequence. This enzyme is useful for synthesizing large amounts of RNA in vivo or in vitro, and is capable of producing a single RNA selectively from a complex mixture of DNAs. The procedure used to obtain a clone of the R7 RNA polymerase gene can be applied to other T7-like phages to obtain clones that produce RNA polymerases having different promoter specificities, different bacterial hosts, or other desirable properties. T7 RNA polymerase is also used in a system for selective, high-level synthesis of RNAs and proteins in suitable host cells.

5 Claims, 6 Drawing Sheets

EXHIBIT S

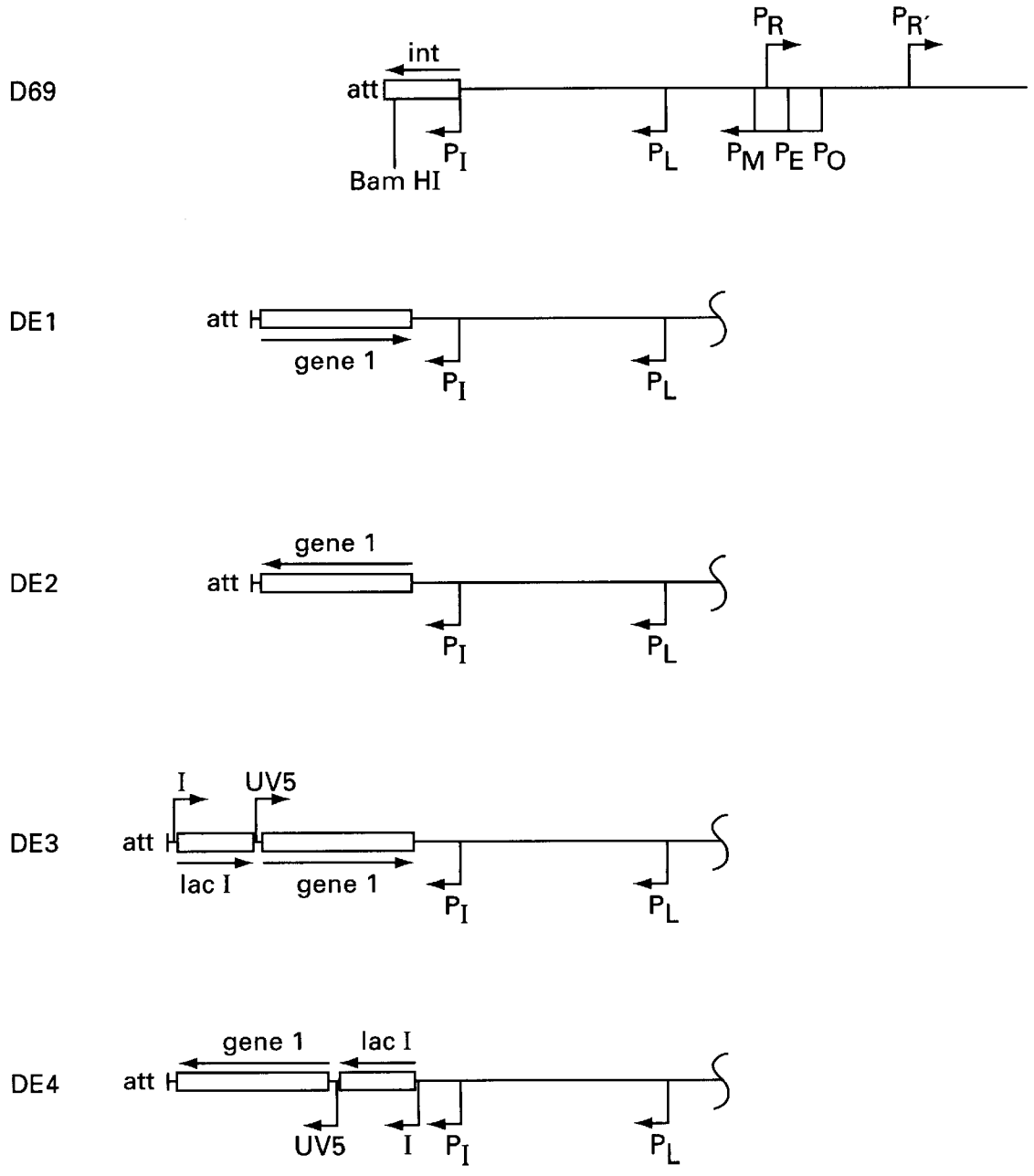

D69
int
att
Bam HI
P_I
P_L
P_R
$P_{R'}$
P_M P_E P_O
DE1
att H
gene 1
P_I
P_L
DE2
gene 1
att H
P_I
P_L
DE3
I
UV5
att H
lac I
gene 1
P_I
P_L
DE4
gene 1
lac I
att H
UV5
I
P_I
P_L

FIGURE 1

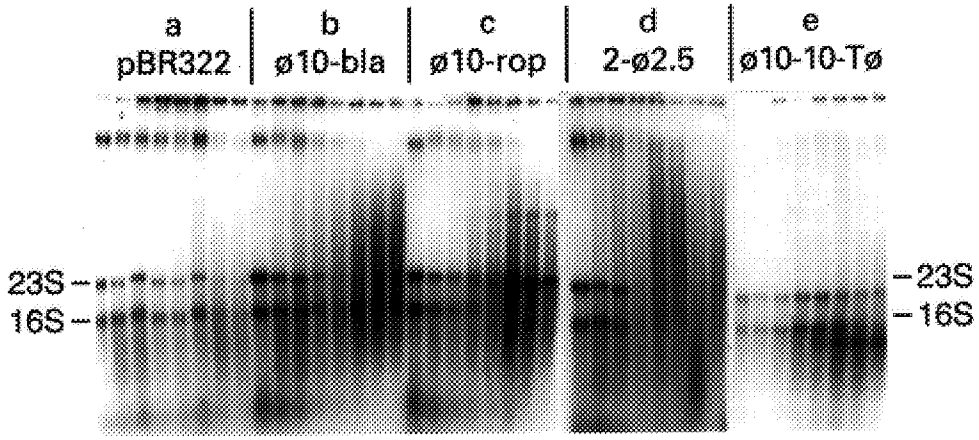


FIGURE 2

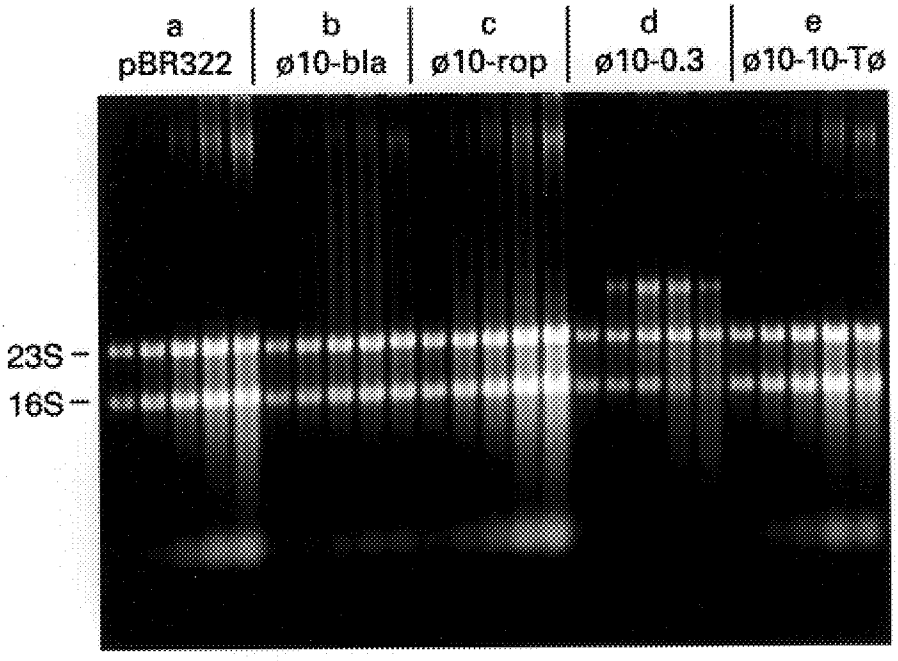


FIGURE 3

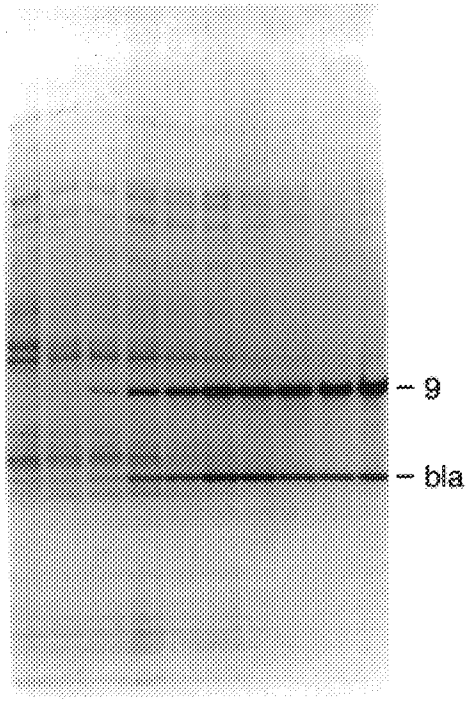


FIGURE 4

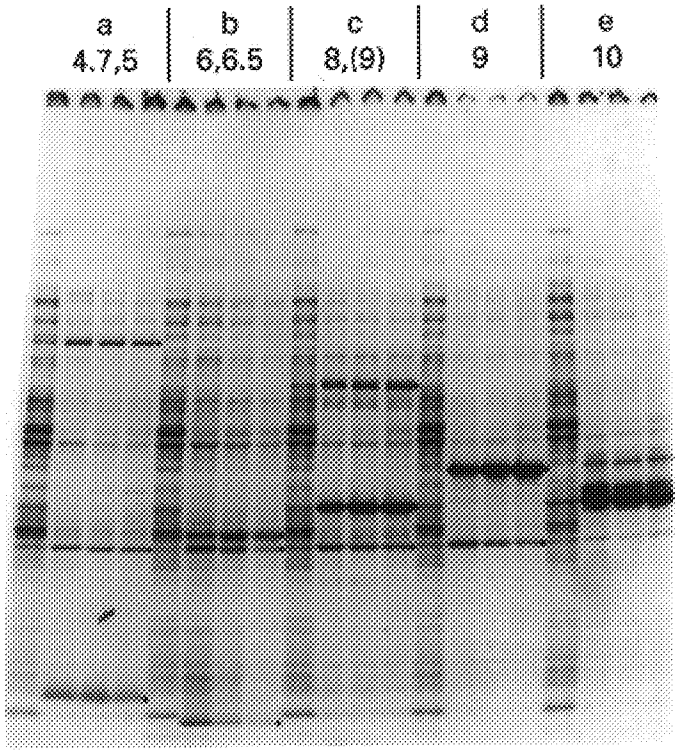


FIGURE 5

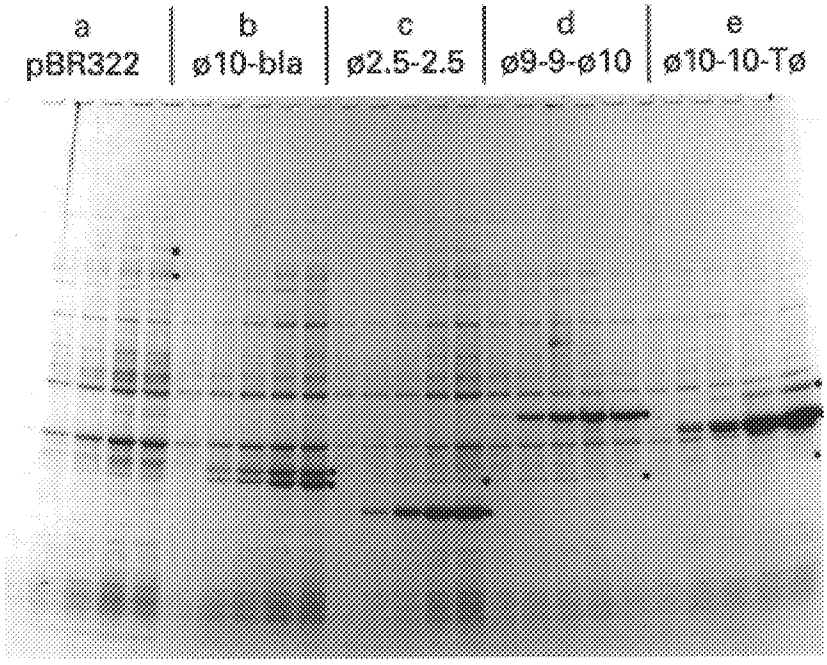


FIGURE 6

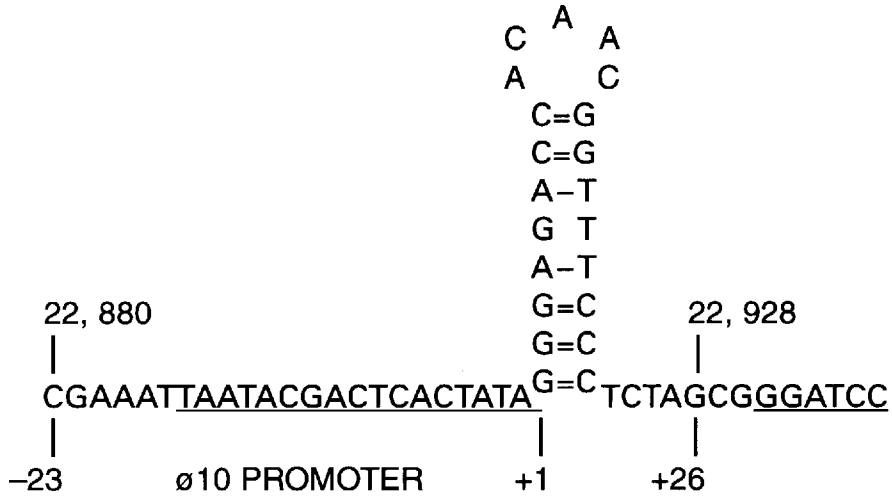


FIGURE 7

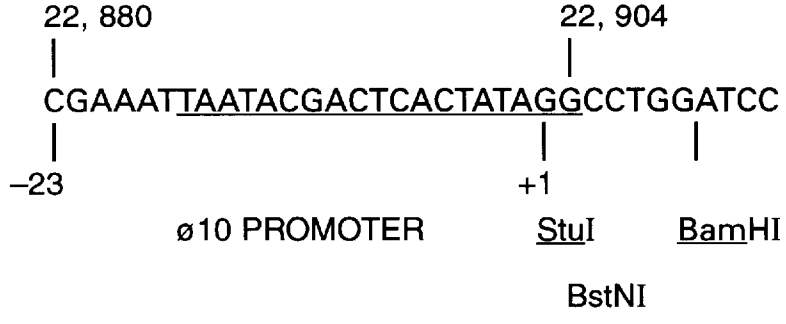


FIGURE 8

FIGURE 9

FIGURE 10

CLONING AND EXPRESSION OF THE GENE FOR BACTERIOPHAGE T7 RNA POLYMERASE

RELATED APPLICATIONS

This is a Continuation of application Ser. No. 08/259,560, filed Jun. 14, 1994, (now U.S. Pat. No. 5,693,489 issued Dec. 2, 1997) which is a Continuation of application Ser. No. 07/874,386, filed Apr. 27, 1992, (now abandoned), which is a Continuation of application Ser. No. 07/542,056, filed Jun. 22, 1990 (now abandoned), which is a Continuation in Part of application Ser. No. 07/002,408, filed Dec. 29, 1986 (now U.S. Pat. No. 4,952,496, issued Aug. 28, 1990), which is a Continuation-In-Part of U.S. application Ser. No. 595,016, filed Mar. 30, 1984 (now abandoned).

This invention was made with Government support under contract number DE-AC02-76CH00016, between the U.S. Department of Energy and Associated Universities, Inc. The Government has certain rights in the invention.

BACKGROUND OF THE INVENTION

Bacteriophage T7 is a virulent bacteriophage that infects *Escherichia coli*. It belongs to a class of bacteriophages that specify relatively simple RNA polymerases that selectively transcribe the DNA of their own bacteriophage but do not transcribe unrelated DNAs [Hausmann, *Current Topics in Microbiology and Immunology,* 75, 77–110 (1976); Korsten, et al., *J. Gen. Virol.,* 43, 57–73 (1979); Towle, et al., *J. Biol. Chem.,* 250, 1723–1733 (1975); Butler and Chamberlin, *J. Biol. Chem.,* 257, 5772–5778 (1982); Chamberlin, et al., *Nature,* 228, 227–231 (1970); Dunn, et al., *Nature New Biology,* 230, 94–96 (1971)]. The T7 bacteriophage has been the subject of extensive scientific inquiry, in part because of its simple yet highly specific RNA polymerase. The genetic organization of T7 and the pattern of gene expression during infection are well understood, and the entire nucleotide sequence of T7 DNA is known (Studier and Dunn, *Cold Spring Harbor Symp. Quant. Biol.,* 47, 999–1007 (1983); Dunn and Studier, *J. Mol. Biol.,* 166, 477–535 (1983); Moffatt, et al., *J. Mol. Biol.,* 173, 265–269 (1984)]. These papers review and provide further references to a considerable body of work on T7 RNA polymerase.

T7 RNA polymerase, the product of T7 gene 1, is a protein produced early in T7 infection; it is a single-chain enzyme with a molecular weight close to 100,000. It appears that the basis for the selectivity of the T7 RNA polymerase is the interaction of the RNA polymerase with a relatively large promoter sequence, a sequence large enough that it is unlikely to be found by chance in any unrelated DNA. In the case of T7, the highly conserved promoter sequence appears to consist of approximately 23 continuous base pairs, which includes the start site for the RNA chain. If exact specification of even as few as 15 of these base pairs were required for initiation of chains, chance occurrence of a functional promoter would be expected less than once in a billion nucleotides of DNA.

The stringent specificity of T7-like RNA polymerases for their own promoter sequences is used by these phages to direct all transcription and replication to their own DNAs during infection. After the phage RNA polymerase is made, other phage gene products inactivate the host RNA polymerase, leaving all transcription in the cell directed by the phage enzyme.

T7 RNA polymerase is very efficient at transcribing DNA from its own promoters, and elongates RNA chains about five times faster than does *E. coli* RNA polymerase [Golomb and Chamberlin, *J. Biol. Chem.,* 249, 2858–2863 (1974)]. Termination signals for T7 RNA polymerase do not seem to occur very frequently, and termination is usually not very efficient [McAllister, et al., *J. Mol. Biol.,* 153, 527–544 (1981)].

Their selectivity, activity, and ability to produce complete transcripts make T7 RNA polymerase and the equivalent RNA polymerases from T7-like phages useful for a variety of purposes. T7 RNA polymerase and SP6 RNA polymerase have been purified from infected cells and have been used to produce RNAs for translation in vitro [Dunn and Studier, *J. Mol. Biol.,* 166, 477–535 (1983)], substrates for splicing [Green, et al., *Cell,* 32, 681–694 (1983)], and hybridization probes [Zinn, et al., *Cell,* 34, 865–879 (1983)]. T7 RNA polymerase made during T7 infection directs the expression of genes under control of T7 promoters in plasmids [Campbell, et al., *PNAS USA,* 75:2276–2280 (1978); Studier and Rosenberg, *J. Mol. Biol,* 153:503–525 (1981); and McAllister, et al., *J. Mol. Biol.,* 153:527–544 (1981)], but these gene products do not accumulate to high levels because of competition from promoters in T7 DNA and because the T7 infection quickly kills the cell. It was anticipated [McAllister, et al., *J. Mol. Biol.,* 153:527–544 (1981)], and the present invention demonstrates, that T7 RNA polymerase would also be useful for directing high-level expression of selected genes in cells. The problem in designing a high-level expression system is how to deliver active T7 RNA polymerase to a cell that contains a T7 promoter.

In the past, phage RNA polymerases like T7 RNA polymerase could be obtained only by infection with the phage from which they derive. The yield of purified RNA polymerase from infected cells is not particularly good, because the enzyme is synthesized for only a few minutes during the infection and does not accumulate to high levels. Nor is infection by these phages an efficient way to direct the transcription of non-phage genes inside the cell, because there is competition from promoters in the phage DNA itself and because the cells lyse within a short time.

Production of active T7 RNA polymerase from the cloned gene is an obvious way to obtain large amounts of enzyme for purification, and to have a source of enzyme that could be introduced into a variety of cells without the disadvantages associated with infection by T7 itself. Presumably recognizing this, other workers attempted to clone the active gene from T7 DNA but were not successful. In one report, Stahl and Zinn [*J. Mol. Biol.,* 148:481–485 (1981)] obtained a clone of the entire gene except for the last nucleotide of the termination codon. However, loss of the termination codon causes additional amino acids to be added to the carboxy terminus, and the protein produced from the clone was inactive.

The present invention discloses a successful process for cloning and expressing the T7 RNA polymerase gene, a process that can also be applied to clone the RNA polymerase genes from other T7-like phages. The same method has subsequently been applied by Tabor and Richardson, *PNAS USA,* 82:1074–1078 (1985) to obtain a different clone of the T7 RNA polymerase gene, and by Morris, et al., *Gene,* 41:193–200 (1986) to obtain a clone of the T3 RNA polymerase gene. Having a clone of the active gene enables the use of it for making large amounts of RNA polymerase for purification, and also enables the use of it to direct sustained, high-level expression of selected genes in the cells. The present invention discloses successful methods for implementing these uses of the cloned gene for T7 RNA polymerase in *E. coli.*

UTILITY STATEMENT

The ability of T7 RNA polymerase and equivalent RNA polymerases from T7-like phages to transcribe selectively any DNA that is linked to an appropriate promoter can serve as the basis for a very specific and efficient production of desired RNAs both in vitro and inside a cell. RNAs produced in vitro can be useful as hybridization probes (for specific medical diagnosis, for example), as mRNAs for in vitro protein synthesis, as substrates for analyzing processing reactions or RNA splicing, or for any purpose requiring a specific RNA. RNAs produced inside the cell can direct the production of specific proteins of interest (antigens for vaccines, hormones, enzymes, or other proteins of medical or commercial value, for example), can form complexes with specific mRNAs to inhibit their translation selectively, in vitro or in the cell, or can be useful for any purpose requiring a specific RNA. Potentially, the selectivity and efficiency of the phage RNA polymerase could make such production very efficient. Furthermore, the unique properties of these phage RNA polymerases may make it possible for them to direct efficient expression of genes that are expressed only inefficiently or not at all by other RNA polymerases. These phage polymerases have the further advantage that it is possible to selectively inhibit the host cell RNA polymerase so that all transcription in the cell will be due to the phage RNA, polymerase.

This invention discloses the first means of cloning the active gene for T7 RNA polymerase, a process that can also be used to clone the RNA polymerases of other T7-like phages. It also discloses a process whereby large amounts of T7 RNA polymerase can be expressed and purified, and processes whereby T7 RNA polymerase from the cloned gene, and potentially the RNA polymerases of any of the T7-like phages, can direct the production of large amounts of RNAs and proteins in cells.

DEPOSIT

Three plasmids of this invention have been deposited in the American Type Culture Collection prior to the filing date of this application and in accordance wish the permanency and accessibility requirements of the U.S. Patent and Trademark Office. Plasmid pAR1151 has ATCC No. 39561; pAR1173 has ATCC No. 39562; and pAR1219 has ATCC No. 39563.

DESCRIPTION OF THE FIGURES

FIG. **1**. Locations and orientations of *E. coli* promoters and inserted genes in D69, DE1, DE2, DE3 and DE4. These phages are also described in Table 1. The DNA between the attachment site (att) and the right end of the mature phage DNA is represented (for a map of the entire lambda DNA molecule See phage 473 of Hendrix et al., *Lambda II*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. 1983). The locations of known *E. coli* promoters and the BamHI cloning site within the int gene of D69, and the locations and orientations of the lacI promoter, the lacI gene, the lacUV5 promoter, and T7 gene 1 in the different derivatives of D69 are represented to scale.

FIG. **2**. Time-course of RNA synthesis after infection of different plasmid-carrying cultures by CE6. Cultures of HMS174 carrying different plasmids were grown in modified B2 medium containing 20 μg ampicillin/ml and infected with approximately 7 infectious particles of CE6/cell, as described in example 7. Samples of culture were labeled for 5 min with $^{32}PO_4$ (25 μCi/ml), and the cells were collected by centrifugation and suspended in an equal volume of 10 mM-sodium phosphate (pH 7.0), 2 mM-Na_3EDTA, 1% (w/v) sodium dodecyl sulfate, 1% (v/v) β-mercaptoethanol, 10% (v/v) glycerol, 0.01% (w/v) bromophenol blue. Each sample was mixed with 1/9 vol. 37% (v/v) formaldehyde, placed in a boiling waterbath for 2 min, and then an amount equivalent to 5 μl of culture was subjected to electrophoresis in a 1% (w/v) agarose gel containing 50 mM-sodium phosphate (pH 7.0), 2 mM-Na_3EDTA, 0.1% sodium dodecyl sulfate, followed by autoradiography. From left to right in each set, lanes represent samples labeled beginning 5 min before, and 0, 5, 10, 15, 30, 45 and 60 min after infection. The plasmids were: (a) pBR322, (b) pAR951, (c) pAR1494, (d) pAR219 and (e) pAR436. Where the plasmid contains any T7 promoters. T7 genes or Tϕ, the order of these is indicated above the lanes, Whenever T7 genes are present, transcription from the T7 promoter is directed counterclockwise, which also directs transcription of the bla gene of the plasmid. When no T7 gene is present, the plasmid mRNA that would be transcribed from ϕ10 is indicated, bla for counterclockwise transcription and rop for clockwise. The positions of 23 S and 16 S ribosomal RNAs are indicated.

FIG. **3**. Accumulation of RNA produced by T7 RNA polymerase. Cultures of BL21(DE3) carrying different plasmids were grown in M9ZB medium containing 20 μg ampicillin/ml, and synthesis of T7 RNA polymerase was induced by adding 0.4 mM-IPTG. The titer of a sample of culture removed before IPTG was added showed that almost all cells in each culture contained plasmid. Cells were collected by centrifugation, resuspended in 1.5 vol. 50 mM-Tris-HCl (pH 6–8), 2 mM-Na_3EDTA, 1% sodium dodecyl sulfate, 1% β-mercaptoethanol, 10% glycerol, 0.025% bromophenol blue, heated for 2 min in a boiling waterbath, and an amount equivalent to 3.3 μl of culture was subjected to electrophoresis in a 1.4% agarose gel in 40 mM-Tris-acetate (pH 8.0), 2 mM-Na_3EDTA. RNA was visualized by ethidium bromide fluorescence. From left to right in each set, lanes represent samples collected immediately before, and 0.5, 1, 2 and 3 h after IPTG was added to the culture. The plasmids were: (a) pBR322, (b) pAR951, (c) pAR1494, (d) pAR946 and (e) pAR436. Genes under control of T7 promoters in the plasmid are indicated above each set, as described in the legend to FIG. **2**. Plasmid pAR946 (set (d)) also contains the RO.3 RNase-III cleavage site. The positions of 23 S and 16 S ribosomal RNAS are indicated.

FIG. **4**. Time-course of protein synthesis after infection of HMS174/pAR441 by CE6. A growing culture of HMS174/pAR441 was infected with approximately 14 infectious particles of CE6/cell, samples were labeled for 5 min with [^{35}S]methionine, and the labeled proteins were analyzed by gel electrophorosis followed by autoradiography. More than 99% of the cells in the culture were infected within 6 min, as indicated by loss of ability to form colonies. From left to right in the Figure, lanes represent samples that were labeled beginning at infection, and 5, 10, 15, 20, 30, 45, 60, 90 and 120 in after infection. The positions of the gene 9 and bla proteins are indicated.

FIG. **5**. Time-course of protein synthesis after infection of different plasmid-carrying cultures by CE6. Cultures of HMS174 carrying different plasmids were grown, infected with approximately 8 to 12 infectious particles of CE6/cell (except 25/cell for pAR1012), libeled for 5 min with [^{35}S] methionine, and the labeled proteins analyzed by gel electrophorosis followed by autoradiography. In each culture, more than 96% of the cells were infected within 6 min, as

indicated by loss of ability to form colonies. From left to right in each set, lanes represent samples labeled immediately before, and 1, 2 and 3 h after infection. The plasmids were: (a) pAR1012, (b) pAR525, (c) pAP213, (d) pAR441 and (e) pAR436. The T7 proteins expressed under direction of T7 promoters in the plasmid are indicated above each set of lanes and marked by dots in the Figure; both the gene 10A and gene 10B proteins are marked. β-Lactamase is also expressed from T7 promoters in all of these plasmids, and its position is also marked in each set. The sizes of these proteins are given in Table 3, except for the fragment of gene 9 protein in set (c), which is predicted to contain 243 amino acid residues.

FIG. **6**. Accumulation of protein directed by T7 RNA polymerase. Cultures of BL21(DE3) carrying the indicated plasmids were grown in ZY medium containing 200 μg ampicillin/ml, and synthesis of T7 RNA polymerase was induced by adding 0.4 mM-IPTG. The titer of samples of culture removed before IPTG was added showed that almost all of the cells had plasmid in each culture except that containing pAR511 (set (c)), where only about 87% of the cells had plasmid. Cells were collected by centrifugation, an amount equivalent to 10 l of culture was subjected to gel electrophoresis, and the proteins were visualized by staining with Coomassie brilliant blue. From left to right in each set, lanes represent samples collected immediately before, and 0.5, 1, 2 and 3 h after IPTG was added to the culture. The plasmids were: (a) pBR322, (b) pAR951, (c) pAR511, (d) pAR441 and (e) pAR436. Genes under control of T7 promoters in the plasmid are indicated above each set, as described in the legend to FIG. **2**. The positions of β-galactosidase (higher) and T7 RNA polymerase (lower), whose synthesis is induced by IPTG, are marked by dots in set (a). The positions of the proteins expressed from the plasmid under control of T7 promoters are marked in the other sets, including β-lactamase in sets (b) to (e) and both the gene 10A and 10B proteins in set (e).

FIG. **7**. Nucleotide sequence of the cloned φ10 promoter (−23 to +26), showing positions of the upstream conserved sequence of the φ10 promoter, the RNA start, a potential stem-and-loop structure in the RNA, and the downstream BamHI cloning site.

FIG. **8**. Nucleotide sequence of the cloned φ10 promoter (−23 to +2), showing the conserved promoter sequence and the downstream linking sequence containing StuI, BstNI, and BamHI sites.

FIG. **9**. Nucleotide sequence from the beginning of the cloned fragment of T7 DNA to the initiation codon of gene 10, showing positions of the φ10 promoter, the RNA start, a potential stem-and-loop structure in the RNA, a unique XbaI site, and the Shine-Dalgarno sequence.

FIG. **10**. Nucleotide and amino acid sequences at the beginning of the gene 10 protein in the plasmids used for expressing coding sequences directly from the initiation codon or as fusions to the gene 10 protein.

DETAILED DESCRIPTION OF THE INVENTION

T7 RNA polymerase is active and efficient, is selective for its own promoters, and will make complete transcripts of almost any DNA that is linked to a T7 promoter. These properties make the enzyme very useful, but in combination they are also responsible for the failure of previous workers to clone the active T7 RNA polymerase gene. We recognized that a specific promoter for T7 RNA polymerase lies just past the end of the coding sequence, and we reasoned that it might be difficult or impossible to clone a DNA fragment that contained both an active polymerase gene and an active promoter. If such a fragment were cloned in a circular plasmid, for example, the polymerase would be able, by initiating at the promoter and transcribing around the plasmid, to direct the synthesis of its own MRNA. This would lead to an autocatalytic increase both in the level of T7 RNA polymerase and in the rate of transcription of the plasmid, which almost certainly would be lethal to the cell. Potentially, a single molecule of active T7 RNA polymerase would be sufficient to trigger this response, so such a construction would be stable only if there were absolutely no expression of the cloned gene by the host RNA polymerase, a situation that is difficult or impossible to achieve.

Cloning of genes in general does not have to contend with the problem of a gene product that can act on one or more sequences that might also be cloned along with the gene, and perhaps that is why previous workers were unable to clone the active T7 RNA polymerase gene. The process we devised to circumvent this problem for the RNA polymerases of T7-like phages and to obtain a clone of the active RNA polymerase gene consists of three steps, which are illustrated in the case of T7 RNA polymerase:

1) Identify the coding sequence for the RNA polymerase in the phage DNA. This can be done by standard techniques, as has been demonstrated for T7 RNA polymerase, and might include isolating and identifying amber mutations in the polymerase gene by their inability to make late RNAs or proteins [Studier, *Science,* 176:367–376 (1972)]; identifying the location of the gene in the phage DNA by testing for ability of mutants to recombine with cloned DNA fragments [Studier and Rosenberg, *J. Mol. Biol.,* 153:503–525 (1981)]; and if necessary by determining the nucleotide sequence [Dunn and Studier, *J. Mol. Biol.,* 166:477–535 (1983); Stahl and Zinn, *J. Mol. Biol.,* 148:481–485 (1981)].
2) Identify the locations of the specific promoters for the phage RNA polymerase. Again in the case of T7, this has been done by transcribing the phage DNA or specific fragments of it with the phage RNA polymerase in vitro or in vivo [for example, Carter, et al., *J. Virol.,* 37:636–642 (1981); McAllister, at al., *J. Mol. Biol.,* 153:527–544 (1981)]. T7-like promoters typically have a conserved nucleocide sequence (23 continuous base pairs in the case of T7), and identification of this sequence would enable promoters to be identified in or near the nucleotide sequence of the RNA polymerase gene.
3) Isolate and clone a DNA fragment that contains the entire coding sequence for the RNA polymerase but no active promoter. The critical step in the entire process is to obtain a DNA fragment containing the coding sequence for an active RNA polymerase but no active promoters for the same RNA polymerase. It is also useful, but may not be in all cases necessary to remove any promoters for the RNA polymerase(s) of the host cell in which the clone will be propagated. The promoters can be removed or inactivated by standard techniques that are widely applied. A procedure we used is given in Example 1, but other commonly used procedures are equally applicable. For example, commonly used techniques for in vitro mutagenesis could be employed to introduce unique cleavage sites for restriction endonucleases at appropriate positions within cloned fragments containing the ends of the gene, the promoters could be removed by cleavage, and

the intact gene without any promoters then reassembled from its separate fragments. In the unexpected event that a promoter for the phage RNA polymerase is located within the coding sequence for the RNA polymerase itself, this promoter would have to be inactivated by the now standard techniques of in vitro mutagenesis to eliminate promoter function without inactivating the gene. Once the nucleotide sequence of an RNA polymerase gene and its promoters is known, it would even be possible to synthesize a gene that will specify the correct amino acid sequence (or one that would make an active enzyme) but will not contain any active promoters.

The above process was used to clone the active gene for T7 RNA polymerase under conditions where its expression would be minimal, in case T7 RNA polymerase should be lethal. The BamHI site of the plasmid pBR322 was used as a cloning site because genes that are lethal if expressed to any significant extent had already been cloned in this site [Studier and Rosenberg, *J. Mol. Biol.,* 153, 503–525 (1981)]. To prevent expression of the cloned gene by the host RNA polymerase, a weak promoter for *E. coli* RNA polymerase that lies just ahead of the gene in the phage DNA was removed.

The above process will also provide active clones of the genes for similar RNA polymerases from other T7-like phages. T7-like phages are widely distributed in nature and infect many different types of bacteria. The genetic organization of all T7-like phages that have been examined has been found to be essentially the same as that of T7. Therefore, it is possible to isolate a DNA fragment containing the coding sequence for the RNA polymerase gene but not external promoters, in a way similar to that described tor T7. Examples of T7-like phages include, but are not limited to *Escherichia coli* phages T3, φI, φII, W31, H, Y, A1, 122, cro, C21, C22, and C23; *Pseudomonas putida* phage gh-1; *Salmonella typhimurium* phage SP6; *Serratia marcescens* phage IV; Citrobacter phage ViIII; and Klebsiella phage No. 11 [Hausmann, *Current Topics in Microbiology and Immunology,* 75, 77–109 (1976); Korsten, et al., *J. Gen. Virol.,* 43, 57–73 (1975); Dunn, et al., *Nature New Biology,* 230, 94–96 (1971); Towle, et al., *J. Biol. Chem.,* 250, 1723–1733 (1975); Butler and Chamberlin, *J. Biol. Chem.,* 257, 5772–5778 (1982)]. The RNA polymerases of other T7-like phages have selectivities for their own promoters that are comparable to the selectivity of T7 RNA polymerase for its promoters, and several different, non-overlapping or partially overlapping promoter specificities are known. Clones of RNA polymerase genes from T7-like phages that are already known (or could be isolated from nature), together with their specific promoters, could be useful for directing the expression of specific genes in bacteria other than *E. coli*. Clones of RNA polymerases having non-overlapping specificities might also be useful for controlling two or more different sets of genes independently in a single cell. Certain T7-like RNA polymerases might also have properties such as temperature, pH, or ionic strength optima, kinetic properties, or stability that would make them particularly useful or desirable for particular applications.

By our preferred design, the clone of the gene for T7 RNA polymerase contains no promoter for *E. coli* RNA polymerase ahead of the gene. This makes it possible to control the expression of the gene by inserting appropriate promoters immediately ahead of it. To facilitate this process, we placed a unique cloning site ahead of the coding sequence so that any inserted promoter plus the coding sequence for T7 RNA polymerase can be removed as a single fragment for insertion into vectors suitable for different cell types (pAR1173, Example 3). In order to be able to make large amounts of T7 RNA polymerase for purification from *E. coli*, we inserted the inducible lacUV5 promoter for *E. coli* RNA polymerase at this site (pAR1219, Example 4). When this promoter is induced, T7 RNA polymerase accumulates to a level of perhaps 10–20% of the cell protein. This enzyme is soluble and active, and a simple purification procedure yields about 10–15 mg of pure protein from 100 ml of induced culture (Example 5 and 6). This is an ample and convenient supply of purified T7 RNA polymerase for most purposes. Of course, much larger amounts of T7 RNA polymerase could be obtained by starting with larger amounts of induced culture, or perhaps by placing a different promoter ahead of the cloned gene.

The cloned gene provides a source of active T7 RNA polymerase inside *E. coli* cells without the disadvantages inherent in T7 infection. Since T7 RNA polymerase is highly selective for its own promoters, which do not occur naturally in *E. coli*, the presence of a T7 promoter in the same cell should direct all transcription by T7 RNA polymerase in the cell to the DNA controlled by that promoter. Potentially, the entire resources of the cell could be directed to the production of selected RNAs and proteins.

In attempting to exploit the clone of gene 1 for this purpose, we initially attempted to place all of the elements needed for this expression system in a single plasmid. Starting with plasmid pAR1219, which contains gene 1 under control of the inducible lacUV5 promoter, we tried to add target genes under control of a T7 promoter. These efforts were unsuccessful, apparently because the basal levels of active T7 RNA polymerase in the uninduced state were such that the plasmids could not be maintained.

T7 RNA polymerase is so active and selective that we considered it might be difficult to find conditions where cloned gene 1 would be shut off to such a degree that the cell could tolerate diverse target genes, many of which might be toxic. To provide a completely general solution to this problem, we decided to remove gene 1 from the cell entirely, and to introduce it only at the time we wished it to be active. In this way, we separate the system into two parts: 1) the source of T7 RNA polymerase from the cloned gene, and 2) the target gene under control of a T7 promoter.

In the two-part configuration, target genes could be cloned in the host cell under control of a T7 promoter without enhancing their expression, since *E. coli* RNA polymerase (and presumably the RNA polymerases of almost any potential host cell) does not initiate transcription at T7 promoters. If target genes are cloned in sites where the host RNA polymerase makes little or no mRNA from them, even genes whose products are very toxic to the cell should be tolerated under control of a T7 promoter. The gene for T7 RNA polymerase could be placed in a virus, but under control of promoters for the host cell or viral RNA polymerase (if any). Upon infection with the virus, active T7 RNA polymerase would be produced and would in turn transcribe the target DNA that is under control of the T7 promoter.

A specific configuration for expression in which the target gene is cloned in the silent orientation in the BamHI site of pBR322 and the T7 RNA polymerase is delivered by a derivative of phage lambda is described in the examples. In this configuration, it is possible to direct essentially all of the resources of the cell to the production of target genes and proteins, and to accumulate large amounts of selected RNAs and proteins. This configuration was used to work out the properties of the T7 expression system, but it is not intended that the process be limited to this specific configuration. We

expect that this process for gene expression will be successful with the target genes cloned in many different types of vectors or in the chromosome of the cell; with T7 RNA polymerase delivered by other types of virus besides lambda; with clones of other T7-like RNA polymerases, alone or in combination, together with target genes controlled by their specific promoters; and in other cell types besides *E. coli*, both prokaryotic and eukaryotic.

A complementary two-part configuration is also possible, in which the target gene is placed under control of a T7 promoter in the virus and the gene for T7 RNA polymerase is maintained in the cell under control of an inducible or constitutive promoter for the host RNA polymerase. In this configuration, it is possible to adjust the level of active T7 RNA polymerase present in the cell before delivery of the target gene by the virus. So far, the levels of expression produced in this configuration have not been as high as in the first process.

Although the properties of the T7 expression system were defined in the first configuration described above, it is also useful to be able to have a system where all parts are resident in the same cell. We have achieved such a unitary process as well. In this configuration a single copy of the gene for T7 RNA polymerase is placed in the chromosome of the cell under control of an inducible promoter for the host cell RNA polymerase and the target gene is placed under control of a T7 promoter in a multi-copy plasmid vector. With only a single copy of the T7 RNA polymerase gene in the cell, the basal activity in the uninduced state is low enough that many but not all target genes can be tolerated in the same cell. When the system is stable enough, the amounts of RNA and protein that can be produced are equally as large as in the first configuration above.

A specific unitary process for expression in which the target gene is cloned in the silent orientation in the BamHI site of pBR322 and the T7 RNA polymerase is placed in the chromosome as a lysogen of a derivative of phage lambda, and where the gene for T7 RNA polymerase is under control of the inducible lacUV5 promoter, is described in the examples. In this configuration, as in the first, it is possible to direct essentially all of the resources of the cell to the production of target genes and proteins, and to accumulate large amounts of selected RNAs and proteins. While this configuration is successful, it is not intended that the process be limited to this specific configuration. We expect that this process for gene expression will be successful with the T7 RNA polymerase gene cloned in the chromosome of the cell under control of many different inducible promoters; with the target genes cloned in many different types of single or multi-copy vectors or in the chromosome; with clones of other T7-like RNA polymerases, alone or in combination, together with target genes controlled by their specific promoters; and in other cell types besides *E. coli*, both prokaryotic and eukaryotic. In particular, since not much T7 RNA polymerase seems to be needed to produce very high levels of expression of target genes, inducible promoters that are very well shut off may be useful in allowing more toxic target genes to be tolerated even if the promoter is not very strong upon induction.

Conditions for optimal expression of target genes in *E. coli* were established using T7 genes under control of T7 promoters (examples 9–11). These examples show clearly that T7 RNA polymerase is capable of focusing the resources of the cell on the production of target proteins under control of a T7 promoter in a plasmid. The extent of the response depends upon the amount of T7 RNA polymerase present, but relatively small amounts generate a large response. The rate of synthesis of the target protein can be much higher than the rate of synthesis of any host protein, and such high rates can continue for three hours or longer. In any case, the large accumulation of RNA makes it seem likely that T7 RNA polymerase transcribing from a T7 promoter in a multicopy plasmid is capable of producing enough mRNA to saturate the protein-synthesizing machinery of *E. coli*.

Plasmid vectors have been constructed (examples 14 and 15) that allow DNA fragments from virtually any source to be placed under control of the strong $\phi 10$ promoter for T7 RNA polymerase, or under control of this promoter plus the protein initiation signals of T7 gene 10, the gene for the major capsid protein of T7. Some of these vectors also contain the transcription termination signal Tϕ from T7 DNA. The control sequences have been cloned so that they can easily be removed from these vectors and inserted in other plasmids, viruses or chromosomes to create a wide variety of other vectors and configurations for T7 RNA polymerase-directed expression of cloned DNA fragments both in vivo and in vitro. And of course, equivalent sets of vectors can also be created from the analagous control elements from other T7-like phages for use with their specific RNA polymerases.

The following examples provide additional elucidation of the plasmids and processes of this invention. Descriptions of the invention can also be found in Davanloo, et al., *PNAS, USA* 81:2035–2039 (1984) and in Studier and Moffatt, *J. Mol. Biol.,* 189:113–130 (1986). These examples are not intended to unduly restrict the invention to the uses described therein. In these examples, the following materials and methods were employed throughout:

1. *E. coli* HMS174[as described in Campbell, et al., *Proc. Natl. Acad. Sci. USA,* 75, 2276–2280 (1978)] or BL21, a Met^+ derivative of B834 [described by Wood, *J. Mol. Biol.,* 16:118–133 (1966)] were used as the host for plasmid strains.
2. Bacteriophage T7, suitable hosts, and techniques for growing and manipulating them were used as described in Studier, *J. Mol. Biol.,* 94, 283–295 (1975); Studier, *J. Mol. Biol.,* 79, 227–236 (1973); and Studier, *Virology,* 39, 562–574 (1969). The nucleotide sequences, locations, and designations of T7 genes and genetic elements are given in Dunn & Studier, *J. Mol. Biol.* 166, 477–535 (1983) and Moffatt, et al., *J. Mol. Biol.* 173, 265–269 (1984).
3. DNA fragments were cloned in the plasmid pBR322, the plasmid described in Bolivar, et al., *Gene,* 2, 95–113 (1977).
4. The preparation and cloning of DNA fragments was by standard techniques as described in *Methods in Enzymology*, Volume 68, (1979), R. Wu, ed., Academic Press, New York, and *Molecular Cloning: A Laboratory Manual* (1982) by Maniatis, et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., and Studier and Rosenberg, *J. Mol. Biol.,* 153, 503–525 (1981).
5. The source of lacI-lacUV5 promoter DNA was the plasmid pMCI described in Calos, *Nature* (*London*), 274, 762–765 (1978). Transcription from the lacUV5 promoter was induced by adding 0.4 mM isopropyl-beta-D-thiogalactoside (IPTG) to growing cultures.
6. Restriction endonucleases and enzymes used in cloning DNA were obtained from New England Biolabs, Bethesda Research Laboratories, or Boehringer Mannheim.

7. Synthetic deoxynucleotide linkers containing a BamHI site or a BglII site were obtained from New England Biolabs.

8. Preparation of the T7 Deletion Mutants. A number of deletions lying to the left or right of gene 1, the gene for T7 RNA polymerase, have been described and characterized [Studier, et al., *J. Mol. Biol.,* 135, 917–937 (1979)]. These deletions apparently arose by genetic crossovers between short repeated sequences. The crossover for D159 is at the sequence A-A-T-G-C-T-G-A, located at nucleotides 975 and 3023 in the nucleotide sequence of T7 DNA, and the crossover for C74 is at the sequence G-T-G-G-C-C-T, located at nucleotides 1458 and 3128 [Dunn and Studier, *J. Mol. Biol.,* 166, 477–535 (1983) and Studier, et al., *J. Mol. Biol.,* 135, 917–937 (1979)]. The likely crossover sequences for LG4 (A-A-T-A-C-G-A-C-T-C-A-C-T-A at 5832 and 7879) and LG26 (G-G-T-A-A-G-A-A at 7165 and 8658) were deduced from restriction mapping and the locations of repeated sequences in this region of T7 DNA. Heteroduplexes between the DNAs of double-deletion mutants D159,LG4 and C74,LG26, when digested with the single-strand-specific nuclease S1, would be expected to produce a fragment extending from nucleotides 3128 to 5845, which would contain the entire coding sequence for T7 RNA polymerase. The double-deletion strains were constructed by conventional genetic crosses, and the presence of both deletions in each double-deletion, strain was confirmed by restriction analysis of the DNA.

9. Growth media. The practice of this invention includes, but is not limited to the following complex 6 growth media: ZB medium (10 g N-Z-amine A/l and 5 g NaCl/l); ZY medium 10 g N-Z-amine A/l, 5 g Bacto yeast extract/l, and 5 g NaCl/l); or M9 medium (1 g NH_4Cl/l, 3 g KH_2PO_4/l, 6 g $Na_2HPO_4$4/l, 4 g glucose/l, and 1 ml 1 m-$MgSO_4$/l); B2 medium (M9 medium in which all but 0.16 mM of the phosphate is replaced by salts and bis-Tris buffer); M9ZB medium (combines M9 and ZB media); M9 maltose and B2 maltose are the equivalent media, in which glucose is replaced by maltose. N-Z-amine A is commercially available from Sheffield Products; Bacto Tryptone and yeast extract are commercially available from Difco. When growing plasmid-containing cells, ampicillin was added to the medium, usually at a concentration of 20 ug/ml, but as high as 200 ug/ml. Dilutions of bacteria or phage for titering were made in ZB medium, and samples were plated by mixing with 2.5 ml of melted top agar [0.7% (w/v) agar in ZB medium], and spreading on plates containing 20 ml of hardened bottom agar (1% agar in ZB medium).

10. Maniatis, et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1982) discloses general procedures for working with lambda; the specific procedures followed for the present invention include: phage stocks were grown in ZY medium. Stock lysates were grown by adding a single plaque 50 μl of a fresh overnight culture of ED8739 to 35 ml of growth medium in a 125 ml flask, and shaking at 37° C. until lysis; larger volumes were grown by adding 10 μl of lysate and 1 ml of cells to 500 ml of medium in a 1 liter flask. Lysates typically contained a few times 10^{10} infective phage particles/ml.

Phage were purified by precipitation with polyethylene glycol, followed by rapid isopycnic banding in CsCl step gradients. All solutions used during purification, including CsCl solutions, contained 10 mM-Tris Hcl (pH 8.0), 10 mM-$MgSO_4$, and 100 μg gelatin/ml to keep the phage intact. The purified phage were stored in the CsCl solution, and dilutions were made in 0.1M-NaCl, 50 mM-Tris Hcl (pH 8.0), 10 mM-$MgSO_4$, and 100 μg gelatin/ml.

11. Cloning vector D69, the cloning vector used in the production of target gene proteins, is a lambda derivative that has imm^{21} and a single BamHI cloning site within the int gene. This vector is described in Mizusawa, et al., *Gene,* 20:317–322 (1982). A mixture of 500 ng of a BamHI digest of D69 DNA and a 5-fold molar excess of a BamHI fragment that contained the gene for T7 RNA polymerase (purified from pAR1151 or pAR1219) was ligated with phage T4 DNA ligase and packaged in a lambda packaging system. About 90% of the resulting plaques contained inserts (of both orientations). Cloning into the BamHI site of D69 interrupts the int gene, whose produce is needed for integration into the chromosome. Therefore, to form lysogens of the D69 derivatives that contain gene 1, int function was provided from a lysogen of heterologous immunity. A drop of lysate was spotted onto a lawn of the helper lysogen, the center of the cleared spot was used to grow a culture, and individual colonies from the culture were tested for the appropriate immunity or for the presence of a functional cloned gene. The genetic composition of D69 and the derivatives used in this process are shown in Table 1 and FIG. **1**.

12. Rates of synthesis and accumulations of RNAs and proteins were analyzed by standard techniques of agarose and polyacrylamide gradient gel electrophoresis, as described in Studier, *J. Mol. Biol.,* 79:237–248 (1973 and *Molecular Cloning: A Laboratory Manual* (1982) by Maniatis, et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

EXAMPLE 1

Isolation of Gene 1 Fragments Lacking T7 Promoters

Search of the nucleotide sequence revealed no restriction sites that would permit convenient isolation of a fragment containing the entire gene 1 coding sequence but not the *E. coli* promoter located ahead of the coding sequence or the T7 promoter immediately following it. Therefore, an appropriate fragment was isolated from heteroduplexes formed between the DNAs of different deletion mutants of T7. The procedure is similar to that employed by Stahl and Zinn, *J. Mol. Biol.,* 148, 481–485 (1981), except that heteroduplexes from overlapping double-deletion mutants were prepared in order to provide an optimal substrate for the single-strand specific endonuclease S1.

Following the techniques described in Studier, *J. Mol. Biol.,* 41, 199–209 (1969), heteroduplexes between the DNA of the double-deletion strains D159,LG4 and C74,LG26 were prepared directly from concentrated stocks of phage particles that had been purified by isopycnic banding in CsCl solution containing small amounts of Tris and EDTA. The stocks in CsCl solution were diluted 50 fold in 30 mM NaOH and left for 10 min. at room temperature to release and denature the DNA. The final DNA concentration was 100 μg/ml. The DNA mixture was neutralized by adding 0.1 volume of 0.5M Tris-Cl, pH 6.8, and the DNA was renatured by incubating 10 min. at 65° C. The renatured DNA was precipitated with ethanol and redissolved in 5 mM Tris-Cl,

pH 6.8. The solution was adjusted to one-third the original volume and to a composition of 0.3M NaCl, 4.5 mM $ZnSO_4$, 30 mM sodium acetate, pH 4.6, for treatment with S1 nuclease. After digestion sufficient to release the double-stranded fragments from the heteroduplexes, the DNA mixture was subjected to electrophoresis on a 1% agarose gel, where the gene 1 fragment migrated as a sharp band well resolved from the other digestion products.

EXAMPLE 2

Construction of Plasmid pAR1151

Because of the possibility that a functional promoter for T7 RNA polymerase might remain in the fragment produced by S1 nuclease digestion of the heteroduplex DNA, we also digested the gene 1 fragments very lightly with BAL-31 exonuclease, to remove a few nucleotides from the ends of the fragment. The fragments generated by S1 alone, or by S1 followed by BAL-31 treatment, were purified by gel electrophoresis and adapted with BamHI linkers (CCGGATCCGG) for insertion into the BamHI site of pBR322. The ligation mixture was used to transform *E. coli* HMS174 to ampicillin resistance, and the transformants were enriched for tetracycline-sensitive clones by treatment with cycloserine.

When individual transformants from several different ligation mixtures were tested by gel electrophoresis of the plasmid DNAs, very few inserts were found, and none contained a complete gene 1. To enrich for plasmids containing an intact gene, plasmid DNA prepared from a mixed population of transformants obtained after cycloserine enrichment was subjected to gel electrophoresis, and DNA was recovered from the region of the gel where plasmids containing complete gene 1 would be expected to migrate. This DNA was then used to transform *E. coli* HMS174, and 48 new tetracylcine-sensitive clones were isolated and analyzed for inserts. Four of these 48 plasmid DNAs appeared to carry an insert of the proper size to contain all of gene 1. Gel electrophoresis of the fragments produced by cutting these plasmid DNAs with HindII and with KpnI produced identical patterns from all four clones, which were essentially the patterns expected if the entire gene had been cloned in the silent orientation. The fragments cloned in this experiment had been treated with BAL-31.

One of the plasmids produced by the above procedure which appeared to carry an insert of the proper size to contain all of gene 1 was pAR1151. To determine exactly where the cloned fragment begins and ends in the nucleotide sequence of T7 DNA, the fragment was released from the plasmid by cutting with BamHI and the nucleotide sequence at each end was determined by the techniques of Maxam and Gilbert (*Methods in Enzymology,* 65, 499–560, Academic Press, New York, 1979). The results showed that, after accounting for the sequence of the BamHI linkers, the fragment that was cloned in pAR1151 begins at nucleotide 3146 of T7 DNA and ends at nucleotide 5840. Because the last nucleotide of the linker at the left end of the fragment happens to be the same as nucleotide 3145 of T7 DNA and the first nucleotide of the linker at the right end the same as nucleotide 5841, the actual limits of T7 DNA sequence are 3145–5841, that is, position 7.88–14.63 in T7 DNA. Therefore, the cloned fragment contains slightly more than the entire coding sequence for T7 RNA polymerase.

Compared with the minimum fragment expected from S1 treatment of the heteroduplex DNA, the cloned fragment of pAR1151 has lost 19 base pairs from the left end and five from the right end, presumably because of the BAL-31 treatment. The first six nucleotides of the natural mRNA are missing, but the fragment does contain 26 base pairs ahead of the AUG initiation codon, and therefore has the entire predicted ribosome-binding and initiation region. The last 46 nucleotides of the natural MRNA are also missing, but the fragment contains 19 base pairs beyond the end of the coding sequence. The cloned fragment contains only the first 11 of the 23 highly conserved base pairs of the promoter for T7 RNA polymerase just past the end of the coding sequence and has no promoter activity for T7 RNA polymerase.

The plasmid pAR1151 was tested to determine if it produced active T7 RNA polymerase. When a cloned fragment of T7 DNA can express a functional gene, T7 amber mutants defective in that gene are usually able to plate with high efficiency on a restrictive host that carries the cloned fragment. In the case of gene 1 amber mutants, active T7 RNA polymerase is absolutely required to produce any increase in plating efficiency above the reversion frequency. We expected that little if any active T7 RNA polymerase would be produced from pAR1151, because the cloned gene is in the silent orientation in the BamHI site of pBR322, where little gene 1 mRNA should be produced. However, we found that gene 1 amber mutants form plaques on HMS174/pAR1151 with the same efficiency as on a host that carries an amber suppressor, although the plaques are somewhat variable in size. Clearly, active T7 RNA polymerase can be made in these cells. The variability of plaque size suggests that the level of active enzyme in the cell may be rather low but high enough that plaques are eventually produced from virtually every infection.

EXAMPLE 3

Construction of Plasmid pAR1173

In order to make large amounts of T7 RNA polymerase from the cloned gene, it was necessary to increase the synthesis of the gene 1 MRNA. To facilitate construction of plasmids in which promoters could be placed ahead of gene 1, a derivative of pAR1151 was first made in which a BglII site was inserted into the BamHI site ahead of gene 1. There are no BglII sites in gene 1 or pBR322, so the inserted BglII site provides a unique cloning site immediately ahead of gene 1. The construction was made by partial digestion of pAR1151 with BamHI, filling in the ends with DNA polymerase, attaching synthetic BglII linkers (CAGATCTG), digesting with BglII, and religating the exposed BglII sites. Because of the sequence of the linkers, a BamHI site should be regenerated to each side of the BglII site. The modified plasmid selected, pAR1173, does not contain a BamHI site between the BlII site and gene 1, presumably due to some imperfection in the cutting and ligation reactions. This is a convenient result because sequences that are inserted in the BglII site will remain with gene 1 as a single fragment upon digestion with BamHI, allowing promoter-gene 1 constructs made in pAR1173 to be moved easily to other vectors.

EXAMPLE 4

Construction of Plasmid pAR1219 and Lambda Derivatives That Carry the T7 RNA Polymerase Gene

One promoter that has been inserted into the BglII site ahead of gene 1 in pAR1173 is the inducible lacUV5 promoter, which remains sensitive to the lac repressor but is

no longer subject to catabolite repression [Silverstone, et al., *PNAS* 66:773–779 (1970)]. The DNA fragment inserted contains the lacI gene as well as the lac UV5 promoter, eliminating the need to rely on the host cell's lac repressor to keep the promoter substantially shut off in the multicopy plasmid. This fragment was obtained from plasmid pMC1 [Calos, *Nature London,* 274:762–765 (1978)]. The 1724 base-pair fragment contains the lacI gene, the lac UV5 promoter region, and the beginning of the lacZ gene isolated from pMC1 by partial digestion with HincII and cloned into the BamHI site of pBR322 using synthetic linkers. The BamHI fragment isolated from this clone was then ligated into the BglII site of pAR1173 ahead of gene 1, taking advantage of the identity of the 4-base extensions produced by BglII and BamHI. A resulting plasmid in which the lac UV5 promoter is directed toward gene 1 is pAR1219. When induced with IPTG, this plasmid produces large amounts of active T7 RNA polymerase.

The active gene for T7 RNA polymerase, alone (pAR1151) or under control of the lac UV5 promoter (pAR1219), was placed in the BamHI site of the lambda cloning vector D69 [described in Mizusawa, et al., cited above). The locations and orientations of *E. coli* promoters and inserted genes in D69, DE1, DE2, DE3, and DE4 are shown in FIG. **1**, and the phages are described in Table 1. All the phages contain both a mutation that eliminates the BamHI site at nucleotide 5505 of lambda DNA, and a deletion of the DNA between the EcoRI sites at nucleotides 21,226 and 26,104. The cloning site in D69 is the BamHI site at nucleotide 27,972 in the int gene. The orientation in which mRNA for T7 RNA polymerase would be transcribed from the p_L and p_I promoters of the phage is designated E, for expressed (in Table 1); the opposite orientation is designated S, for silent. The immunity region between nucleotides 34,379 and 38,617 either has been replaced by the immunity region of phage 21 or is the lambda immunity region having the cI857 and ind1 mutations. The nin5 deletion (Δ) removes nucleotides 40,502, to 43,307. In the orientation represented by DE1, no *E. coli* promoters that would direct transcription of gene 1 mRNA are known to be present in D69 DNA; DE3 contains gene 1 in the same orientation but under control of the lacUV5 promoter. In the opposite orientation, represented by DE2, the p_L and p_I promoters of D69 can direct transcription of gene 1 mRNA; DE4 contains gene 1 in this orientation but is also under control of the lac UV5 promoter. All four of these strains grow well and produce high-titer lysates, and all have been obtained as lysogens of IMS174 and BL21. Apparently, the presence of the T7 RNA polymerase gene, whether expressed or not, has little effect on lambda growth or lysogeny in the usual hosts.

Addional derivatives carrying gene 1 were obtained by crossing DE2 (in which gene 1 is under control of p_L and p_I) with lambda cI857ind1Sam7. The cI857 mutation would make expression from the p_L promoter temperature-inducible, and the Sam7 mutation would prevent lysis of the infected or induced cells. Three recombinants selected for further use all appear to have retained the left arm of DE2 and to have lost the nin5 deletion; DE6 retains imm^{21} and has acquired the Sam7 mutation; CE2 has acquired imm^{λ} (cI857ind1); and CE6has acquired both imm^{λ}(cI857ind1) and Sam7.

TABLE 1

Derivatives of the Lambda Cloning Vector D69 that Carry T7 Gene 1

Phage	Fragment cloned into int	Orientation	Immunity	nin5	Sam7
D69	None		21	Δ	+
DE1	T7 gene 1	S	21	Δ	+
DE2	T7 gene 1	E	21	Δ	+
DE3	lacUV5-gene 1	S	21	Δ	+
DE4	lacUV5-gene 1	E	21	Δ	+
DE6	T7 gene 1	E	21	+	am
CE2	T7 gene 1	E	λcI857	+	+
CE6	T7 gene 1	E	λcI857	+	am

D69 is described by Mizusawa & Ward (1982). All phages in the Table contain both a mutation that eliminates the BamHI site at nucleotide 5505 of lambda DNA and a deletion of the DNA between the EcoRI sites at nucleotides 21,226 and 26,104. The cloning site in D69 is the BamHI site at nucleotide 27,972 in the int gene. The orientation in which mRNA for T7 RNA polymerase would be transcribed from the p_L and p_I promoters of the phage (see FIG. 1) is designated E (for expressed); the opposite orientation is designated S (for silent). The immunity region between nucleotides 34,379 and 38,617 either has been replaced by the immunity region of phage 21 or is the lambda immunity region having the cI857 and indl mutations. The nin5 deletion (Δ) removes nucleotides 40,502 to 43,307.

EXAMPLE 5

Induction of T7 RNA Polymerase from pAR1219

When cultures of HMS174/pAR1219 or BL21/pAR1219 were induced with IPTG, T7 RNA polymerase was produced at a rapid rate and accumulated to levels such that it was the major protein of the cell. Enzyme activity, as assayed in crude extracts, increased along with accumulation of the protein. The rate of increase of turbidity of the culture was almost as high for the induced culture as for a parallel uninduced culture, and colonies could form on plates containing inducer. This suggests that T7 RNA polymerase itself is not very toxic to the cells.

EXAMPLE 6

Purification of T7 RNA Polymerase from BL21/pAR1219

T7 RNA polymerase was isolated from induced cultures of BL21/pAR1219 growing in a shaking flask at 37° C. in a relatively rich medium [tryptone broth plus M9, Studier, *Virology,* 39, 562–574 (1969)]. IPTG was added to a final concentration of 0.4 mM when the cells reached a few times 10^8 per ml and shaking at 37° C. was continued another 4 hrs. These cells contain perhaps 10–20% of the total protein as T7 RNA polymerase, virtually all of it soluble and active.

T7 RNA polymerase can be purified from the induced cells by conventional techniques, for example, Chamberlin, et al., *Nature,* 228, 227–231 (1970), and typically about 10–15 mg of essentially pure protein can be obtained from 100 ml of culture (about 1 g of packed cells). BL21 was used as the host cell because we have found it to lack a protease found in HMS174 (and in many laboratory strains of *E. coli*) that can nick the T7 RNA polymerase protein during purification. However, essentially intact T7 RNA polymerase can be purified from protease-containing strains if the protease is removed at an early stage of the purification.

The purified enzyme is very active and specific in transcribing DNA that has a promoter for T7 RNA polymerase, as demonstrated by electrophoretic analysis of the DNA template and the RNAs produced from it. Almost all of the

RNA produced when purified enzyme was used to transcribe a mixture of two DNA fragments, one of which contained a promoter for T7 RNA polymerase and the other of which did not, was of the size expected if it had started at the promoter and ended at the end of the fragment. This indicates that the transcription was specific and the enzyme was free of RNase activity. At a high enzyme concentration all of the precursors were incorporated into RNA in 15 minutes; at a lower enzyme concentration, incorporation continued for at least an hour. In both cases, the DNA fragments remained intact, indicating that the enzyme was free of DNase activity.

EXAMPLE 7

T7 RNA Polymerase Provided By Infection

Each of the D69 derivatives listed in Table 1, except DE1, produces enough T7 RNA polymerase during infection to generate high-level expression of target genes cloned in a plasmid under control of a T7 promoter. The following procedure, using CE6as the phage, produces efficient infection and high-level expression of the target gene protein.

Cultures are grown at 37° C., usually in a shaking incubator or waterbath, and plasmid-containing cells are grown in the presence of ampicillin, usually at 20 μg/ml. In order to obtain efficient infection (as measured by loss of colony-forming units), cultures are grown in the presence of maltose (and no glucose) to induce the lambda receptor. When the multiplicity of infection is higher than about 20 infectious phage particles per cell, the overall rate of protein synthesis after infection is drastically reduced and little protein is produced from the target gene. Protein synthesis was completely abolished in five different cultures after infection by 56 phage particles per cell, and almost completely abolished in three of them after infection by 28 particles per cell. All of the cultures retained active protein synthesis at multiplicities of 7 and 14, and achieved high rates of synthesis of proteins whose genes are transcribed by T7 RNA polymerase. Inhibition of protein synthesis at high multiplicities is complete within five minutes after adding the phage. Because of this effect, optimal expression of target genes is achieved at multiplicities of 5 to 10—high enough to infect almost every cell, but low enough that protein synthesis is not inhibited.

If cells are grown in M9 maltose, rates of protein synthesis can be measured by pulse-labeling with [^{35}S] methionine (FIG. **5**), and if grown in B2 maltose, rates of RNA synthesis can be measured by pulse-labeling with $^{32}PO_4$ (FIG. **2**). When the absorbance at 500 nm (A_{600}) of an actively growing culture reaches 0.3, glucose is added to give a concentration of 4 mg/ml, and the culture is grown for an additional 1 to 2 hours, during which the A_{600} reaches 0.6 to 1, and the cell concentration is typically 5×10^8/ml. At this point, $MgSO_4$ is added to a final concentration of 10 mM, and purified CE6phage is added to a final concentration of 2×10^9/ml (0.001 volume of a stock having an A_{260} of 6). The multiplicity of infection is typically around 7, and the efficiency of infection is typically greater than 95%, as measured by loss of colony-forming ability. Addition of glucose and 10 mM $MgSO_4$ to the medium is not necessary, but seems to give slightly better production of protein from the target genes. Cells are usually harvested 3 hours after infection, enough time for substantial accumulation of target protein, but not enough time for uninfected cells to overgrow the culture.

EXAMPLE 8

T7 RNA Polymerase Provided by Induction of a Chromosomal Copy of the Gene

The lysogen BL21 (DE3) contains a single copy of the gene for T7 RNA polymerase in the chromosome, under the control of the inducible lacUV5 promoter. Some T7 RNA polymerase is produced from the prophage even in the absence of added inducer, so plasmids containing toxic target genes can be difficult or impossible to maintain in BL21(DE3). However, when the plasmid can be maintained, addition of isopropyl-β-D-thiogalactopyranoside (IPTG) induces the lacUV5 promoter to produce T7 RNA polymerase, which in turn initiates high-level expression of the target gene in the plasmid. The level of expression is usually comparable with that found in Example 7 (by infection).

Because toxicity of the target gene can lead to loss of the plasmid or the accumulation of non-functional mutants, cultures are sampled just before induction and titered for viable cells on plates without additives, or in the presence of 0.5 mg of ampicillin, 2.5 mol of IPTG, or both, added to the top agar. BL21(DE3) carrying a plasmid without a T7 promoter will form colonies on each of these plates, as will non-functional mutants that retain plasmid. In a typical culture useful for producing target proteins, almost all cells will form colonies on plates without additives or containing only amicillin. Less than 2% of the cells will form a colony on plates containing only IPTG, and less than 0.01% will form a colony on plates containing both ampicillin and IPTG.

Cultures have been grown in M9, M9ZB, ZY medium, or ZY medium supplemented with 0.4% (w/v) glucose. Induction is with 0.4 mM IPTG, when the culture reaches an A_{600} of 1 (corresponding to about 5×10^8 to 10^9 viable cells per ml). Cells are typically harvested 3 hours after induction, after substantial accumulation of target protein is achieved, but before the culture can be overgrown with unproductive cells or cells that have lost their plasmid.

EXAMPLE 9

Plasmids Used for Working Out the Properties of the T7 Expression System

Plasmids used to establish the utility of T7 RNA polymerase for directing the transcription and translation of cloned DNA in *E. coli* are listed in Table 2.

TABLE 2

Plasmids	Genetic elements inserted
pBR322	
pAR951	φ10
pAR1494	φ10 (clockwise)
pAR946	φ10-R0.3-0.3
pAR219	2-φ2.5
pAR511	φ2.5-2.5
pAR1012	R4.7-φ4.7-4.7-5
pAR525	6-6.3-φ6.5-R6.5-6.5
pAR213	8-φ9-(9) (243aa)
pAR441	φ9-9-φ10
pAR436	φ10-10-Tφ

All inserts are in the BamHI site of pBR322 and all, except in pAR1494, are oriented so that transcription from T7 promoters proceeds counterclockwise in the conventional representation of pBR322. The order of T7 genetic

elements in the inserted fragments is given, including T7 promoters (ϕ2.5, ϕ4.7, ϕ6.5, ϕ9 or ϕ10), intact T7 genes (0.3, 2, 2.5, 4.7, 5, 6, 6.3, 6.5, 8, 9 or 10), RNase III cleavage sites (R0.3, R4.7, R6.5), and the Tϕ transcription terminator. The sizes and functions of proteins specified by the plasmids are given in Table 3. Plasmid pAR213 carries all of gene 8 and a large fragment of gene 9, predicted to direct a T7-pBR322 fusion protein that is 243 amino acid residues (aa) long.

TABLE 3

Sizes and Functions of Proteins Specified by Plasmids

Gene	No. of amino acids	M_r	Function
A. For pBR322 proteins			
bla precursor	285	31,393	
bla processed	263	28,899	Beta-lactamase
rop	63	7226	Control of replication
B. For T7 proteins			
0.3	116	13,678	Anti-restriction
2	63	7043	Anti-*E. coli* RNA polymerase
2.5	231	25,562	DNA-binding
4.7	135	15,208	Unknown
5	704	79,692	DNA polymerase
6	347	39,995	Exonuclease
6.3	37	4088	Unknown
6.5	84	9474	Unknown
8	535	58,989	Head-tail junction
9	306	33,766	Head assembly
10A	344	36,414	Major head protein
10B	397	41,800	Minor head protein

It is assumed that the bla precursor does not retain the initial methionine residue but that the rop protein does. The gene 10B protein of T7 is produced by frameshifting during translation of the 10A mRNA. The relative mobilities of T7 proteins in gel electrophoresis in the presence of sodium dodecyl sulfate do not always correspond to their relative molecular weights.

EXAMPLE 10

Production of RNA in *E. coli*

T7 RNA polymerase can direct very high level synthesis of RNA, using T7 promoters in plasmids in *E. coli*. Rates of synthesis and levels of accumulation can be comparable to those for ribosomal RNAs (FIGS. **2** and **3**). Transcription from the T7 promoter can also interfere with RNA synthesis by *E. coli* RNA polymerase (FIGS. **2***b* and **2***c*). If no transcription termination signals for T7 RNA polymerase are in the plasmid, transcription can continue around the entire plasmid to produce heterogeneous RNAs larger than plasmid length (FIGS. **2***b*, **2***c* and **2***2d*, and FIGS. **3***b* and **3***c*). Where the RNA contains a single efficient RNase III cleavage site, the large heterogeneous RNAs are cut at this site and accumulate as a discrete band of the length expected for transcription once around the plasmid (FIG. **3***d*). Within an hour after induction, the amount of RNA accumulated in this band can approach that found in the ribosomal RNA bands (FIG. **3***d*). Where the plasmid DNA contains the transcription termination signal Tϕ, transcription terminates efficiently to produce RNA of the expected length (FIGS. **2***e* and **3***e*). The RNAs appear to be relatively stable, and it seems likely that enough RNA accumulates to saturate the translation apparatus of *E. coli*.

Once active T7 RNA polymerase is present in the cell, transcription by the host cell RNA polymerase can be selectively inhibited. In the case of *E. coli*, this can be done by adding inhibitors such as rifampicin, which inhibits *E. coli* RNA polymerase but not T7 RNA polymerase, or by producing an inhibitor such as the gene 2 protein of T7 (FIG. **3***d*), which binds to *E. coli* RNA polymerase and inactivates it. This can be useful in applications where it is desirable or necessary that the transcripts from the T7 promoter be the only ones being produced in the cell.

Example 11

Time Course and Rate of Target Protein Synthesis

T7 gene 9 protein is efficiently synthesized in *E. coli*, and plasmid pAR441 was used to analyze optimum conditions for protein synthesis directed by the T7 expression system. The -lactamase mRNA is also transcribed by T7 RNA polymerase in this plasmid. Upon CE6infection of HMS174/pAR441, synthesis of gene 9 protein and β-lactamase begins to increase 10 to 15 minutes after infection, and by 15 to 20 minutes after infection these proteins are the most rapidly synthesized in the cell. Their rate of synthesis continues to increase until at least 30 minutes after infection and remains at a very high level for at least another 90 minutes. During this period, the rate of synthesis of host proteins gradually declines, and there is little evidence for synthesis of lambda proteins. Synthesis of T7 RNA polymerase can be detected, but remains at a relatively low level; apparently, large amounts of T7 RNA polymerase are not needed in order to direct most of the protein synthetic capacity of the cell to the production of target proteins. The rate of synthesis of gene 9 protein is considerably higher than that of β-lactamase, even though there is every indication that the two mRNAs are produced in comparable amounts and are comparably stable.

EXAMPLE 12

Target Protein Synthesis under Different Conditions for Delivery of T7 RNA Polymerase

Experiments similar to those shown in Example 11 have examined stimulation of synthesis of gene 9 protein and B-lactamase after infection of HMS174/pAR441 by D69, DE1, DE3, and DE4 (see FIG. **1** for the location of gene 1 relative to *E. coli* promoters in these phages). As expected, no stimulation of target protein synthesis is observed after infection by D69, which can not supply any T7 RNA polymerase; nor is any stimulation observed after infection by DE1, a result indicating that little if any transcription of gene 1 occurs during infection by DE1. Stimulation by DE2, on the other hand, is comparable to that by CE6, as is stimulation by DE3 or DE4 in the presence of IPTG. In the absence of inducer, DE3 stimulates some target protein synthesis, but not as much as in the presence of inducer; apparently, the infecting copies of the lacUV5 promoter are only partially repressed. Likewise, DE4 stimulates considerably less target protein synthesis in the absence of inducer than in its presence or than is stimulated by DE2 (which contains no lacUV5 promoter); apparently, repression of the lacUV5 promoter in DE4 blocks transcription initiated at p_L and/or p_I. Finally, when DE2 infects an immune host, target protein synthesis is delayed by a few minutes and may not reach quite as high a level as in a sensitive host, but very substantial production of target protein occurs none the less. In an immune host, p_L should be repressed; perhaps more DNAs enter the cell than can be repressed by the cI protein present at the time of infection, or perhaps transcription from the p_I promoter, which is known to function at a low level in an immune host, is sufficient to produce this response.

Transcription of the target plasmid by T7 RNA polymerase is so active that it interferes with expression of lambda proteins during infection. As a result, the lambda infection is aborted and the cells do not lyse. Therefore, use of inhibitors of *E. coli* RNA polymerase, or infecting an immune cell to prevent lambda gene expression, or use of lambda mutants that cannot lyse the cell, are generally not needed nor useful for enhancing or extending target gene expression.

Comparably high levels of expression are obtained whether T7 RNA polymerase is delivered by a phage carrier or is induced from a chromosomal copy of the gene. Making a lysogen of DE3 is a convenient way to place a single copy of gene 1 into the chromosome under control of the lacUV5 promoter. Such lysogens are very stable because gene 1 is inserted into the int gene, whose produce is needed for excising the prophage from the chromosome. The basal level of T7 RNA polymerase in the uninduced single-copy lysogen is considerably lower than when the cell contains the multi-copy plasmids pAR1151 or pAR1219, as measured by ability to plate 4107, a deletion mutant of T7 that completely lacks gene 1. This low basal activity allows a wide variety of target genes under control of a T7 promoter to be tolerated in the cell. Target genes that are too toxic to be established in a DE3 lysogen can still be expressed by infection with a gene 1-carrying lambda derivative.

EXAMPLE 13

Production of different target proteins in *E. coli*

A variety of different T7 proteins having different degrees of toxicity to the host cell have been expressed in the T7 expression system, where T7 RNA polymerase has been delivered by infection with a lambda derivative or produced by induction from a chromosomal copy of gene 1. Protein patterns illustrating rates of target protein synthesis at different times after synthesis of T7 RNA polymerase was initiated are shown in FIG. **5**, and patterns illustrating the accumulation of target proteins are shown in FIG. **6**. Typically, the target proteins are produced at substantial rates, and synthesis of host proteins is greatly reduced. This shift in protein synthesis reflects the shift in mRNA population in favor of transcripts produced by T7 RNA polymerase, and the relatively high translational efficiency of the T7 mRNAs. High rates of synthesis of the target proteins can continue for at least three hours (as shown in FIG. **5**) and have been observed to continue for at least seven hours in some of the cases we have examined. Target proteins can accumulate to become a substantial fraction of the total cell protein, perhaps 50% or more, within a few hours (FIG. **6**). The relative rates of synthesis and accumulation of individual T7 proteins expressed in the T7 expression system differ considerably, and generally seem to parallel the relative rates of synthesis during T7 infection. It seems likely that the T7 expression system produces saturating amounts of mRNA, and that the relative rates of protein synthesis reflect mainly the relative efficiencies of translation of the individual T7 mRNAs.

EXAMPLE 14

Vectors for Transcribing Target DNAs

T7 RNA polymerase is potentially capable of transcribing any DNA that is placed under control of a T7 promoter. Table 4 lists a set of plasmid vectors that contain a strong promoter for T7 RNA polymerase followed by one or more unique cloning sites. In pAR2529, the cloning site is followed by Tφ, the transcription terminator. These vectors are all derived from pBP322, and DNA from many sources can be cloned in them so that it can be transcribed by T7 RNA polymerase in vitro or in vivo. The DNA fragments containing the T7 promoter or T7 promoter and terminator have been designed so that they can easily be removed from certain of the vectors and used to create new expression sites in other plasmids, viruses, or chromosomes. The practitioner of the art will recognize that a wide range of different expression vectors could be created using these or similar DNA fragments that contain a T7 promoter or T7 promoter and terminator. Detailed description of some of the more useful vectors is disclosed in Rosenberg, et al. *Gene,* 56, 125–135 (1987) and follows.

TABLE 4

Vectors for transcribing cloned DNAs

Vector	Isolation No.	Upstream site	Promoter	Cloning sites	Downstream Elements
	pAR1959	BamHI	φ10	BamHI	
pET-1	pAR2019		φ10	BamHI	
pET-2	pAR2305	BglII	φ10	BamHI	
pET-3	pAR2529	BglII	φ10	BamHI	Tφ BglII
pET-5	pAR2192		φ10	BamHI-EcoRI	Δ in pBR322
pET-6	pAR2369	BamHI	φ10	StuI-BamHI	
pET-7	pAR2463	BglII	φ10	StuI-BamHI	

Plasmid pAR1959 was constructed by inserting a TaqI-XbaI fragment of T7 DNA (nucleotides 22,880–22,928) containing the φ10 promoter for T7 RNA polymerase into the BamHI site of pBR322 by using the synthetic linker CGGGATCCCG. The fragment extends from nucleotides −23 to +26 relative to the start of the RNA and is oriented so that transcription from the φ10 promoter is directed counterclockwise, opposite to transcription from the tetracycline promoter. Counterclockwise transcription through this region by *E. coli* RNA polymerase is low enough to permit relatively toxic genes to be cloned under control of the T7 promoter. T7 RNA polymerase initiates active and selective transcription at the φ10 promoter both in vivo and in vitro.

To increase the usefulness and versatility of plasmids containing this φ10 promoter fragment, the following derivatives were constructed:

pAR2019, in which the upstream BamHI site of pAR1959 has been removed by opening, filling in, and re-ligating. This leaves the downstream BamHI site as a unique cloning site.

pAR2305, in which the upstream BamHI site of pAR1959 has been converted to a BglII site by opening, filling in, adding the linker GAGATCTC, cutting with BglII, and re-ligating. Both the upstream BglII site and the downstream BamHI site are unique in this plasmid. New φ10 vectors can be created by moving the BglII-BamHI fragment to a unique BamHI site.

pAR2529, in which the transcription terminator for T7 RNA polymerase, Tφ, has been added just downstream of the BamHI cloning site of pAR2305. A fragment of T7 DNA containing T (nucleotides 24,106–24,228, where transcription terminates at nucleotide 24,209) was joined to the BamHI cloning site of pAR2305 through the sequence [GGATCC]GG-Tφ-CCGGATCGAGATCTCGATCC, where the final C is nucleotide 375 in the BamHI site of pBR322. The downstream linker contains a BglII site, so the entire φ10-BamHI-T fragment can be removed from this plasmid as a BglII fragment for transfer to other vectors.

pAR2192, in which the DNA between the BamHI and EcoRI sites of pAR2019 (originally from pBR322) has been deleted to leave the BamHI and EcoRI sites in the sequence GGATCCGTTAAC. This deletion removes the tetracycline promoter, so transcription by *E. coli* RNA polymerase across these cloning sites should be relatively low in both directions. If a fragment cloned into the BamHI site has no EcoRI sites, cutting the plasmid at the EcoRI site will cause the RNA made by purified T7 RNA polymerase to end only a few nucleotides past the end of the cloned fragment.

In all of the above plasmids, the RNA transcribed from a cloned DNA will begin with 26 nucleotides of T7 sequence and a CG from the linker sequence before the first G of the BamHI cloning site. The first 21 nucleotides of the RNA could potentially fold into a relatively stable 8 base-pair stem-and-loop structure (FIG. **7**).

In order to be able to make RNA with only two nucleotides preceding the transcript of a cloned DNA, we took advantage of the fact that nucleotides −1 to +2 of the ϕ10 promoter sequence are AGG, which is half of a StuI site (AGG'CCT). The cloned sequence downstream of +2 in pAR1959 and pAR2305 was replaced by the remaining half of the StuI site followed immediately by a BamHI site in the sequence CCTGGATCC to create pAR2369 and pAR2463 (FIG. **8**). The StuI site is unique in these plasmids, and can be used to place any DNA fragment at position +3 by blunt-end ligation.

In spite of the replacement of nucleotides +3 to +6 of the conserved sequence of the ϕ10 promoter, T7 RNA polymerase initiates RNA chains efficiently at this promoter in both plasmids. We expect that most or all DNA fragments inserted at the StuI site will be transcribed by T7 RNA polymerase to produce RNAs having only an additional GG at their **5'** end. Cutting at the downstream BamHI site can also be used to end transcription only a few nucleotides past the end of the cloned fragment.

The sequence CCTGG found in the StuI-BamHI sequence of pAR2369 and pAR2463 is a cleavage site for BstNI and also a methylation site for the dcm methylase of *E. coli*. Methylation at this site prevents cleavage by StuI, so plasmid must be prepared from a dcm⁻ strain if it is to be cut efficiently at the StuI site. BL21 is a B strain of *E. coli* and as such is dcm⁻.

EXAMPLE 15

Vectors for Transcribing and Translating DNAs in *E. coli*

Different mRNAs are translated with widely different efficiencies in *E. coli*. The factors affecting efficiency of translation are not completely understood, but coding sequences usually start with AUG, and the mRNA usually contains upstream sequences (the Shine-Dalgarno sequences) thought to be important in binding the mRNA to the ribosome and initiating translation. The structure of the mRNA is also thought to be important.

The major capsid protein of T7, specified by gene 10, is made very efficiently during T7 infection, much more rapidly than any host protein. The mRNA initiated at the ϕ10 promoter contains 63 nucleotides ahead of the gene 10 initiation codon (FIG. **9**), and this leader sequence seems likely to be responsible at least in part for the efficiency of translation of the gene 10 mRNA. The mRNA begins with 21 nucleotides that could form a relatively stable stem-and-loop structure, and continues with an apparently unstructured, AT-rich region that contains a good Shine-Dalgarno sequence ahead of the initiation codon. The sequence containing the initiation codon forms part of an NdeI cleavage site in the DNA (CATATG). Specific cleavage of the DNA at this site can be used to fuse coding sequences directly to the initiation codon of gene 10. An NdeI cleavage site is particularly useful for this purpose because the initiation codon for any protein can potentially be made to be part of an NdeI cleavage site in its DNA by changing only upstream, noncoding nucleotides, and without changing any coding sequences for the protein. Fusing coding sequences to the leader sequence in such a way could place the initiation codon for any protein at a site known to be efficiently translated in *E. coli*.

We have constructed vectors containing the 10 promoter and gene 10 protein initiation region to a point just past the 11th codon for the gene 10 protein. Besides the NdeI cleavage site at the initiation codon, the sequence contains an NheI cleavage site at the second and third codons. In addition, we have placed BamHI cleavage sites after the 11th codon, using linker oligonucleotides of three different lengths. Such a set of three vectors makes it relatively easy to fuse any coding sequence after the 11th codon of the gene 10 protein so that it will be in the correct reading frame to be translated from the gene 10 initiation site. In one set of vectors, the BamHI cloning sites are followed by Tϕ, the transcription terminator. The DNA fragments containing the control sequences have been designed so that they can easily be removed from certain of the vectors and used to create transcription-translation sites in other plasmids, viruses or chromosomes. The practitioner of the art will recognize that a wide range of different expression vectors could be created using these or similar DNA fragments that contain a T7 promoter and a translation initiation site, with or without a transcription terminator. Detailed description of some of the more useful vectors is disclosed in Rosenberg, et al. *Gene,* 56, 125–135 (1987) and follows.

TABLE 5

Vectors for Transcribing and Translating Cloned DNAs

Vector	Isolation No.	Upstream site	Expression signals	Fusion cloning site	Open frame	Down-stream
	pAR2075	BamHI	ϕ10-s10	BamHI (8mer)	ATC	
	pAR2078	BamHI	ϕ10-s10	BamHI (10mer)	GAT	
	pAR2084	BamHI	ϕ10-s10	BamHI (12mer)	GGA	
pET-1c	pAR2120		ϕ10-s10	BamHI (8mer)	ATC	
pET-1b	pAR2093		ϕ10-s10	BamHI (10mer)	GAT	
pET-1a	pAR2098		ϕ10-s10	BamHI (12mer)	GGA	
pET-2c	pAR2156	BglII	ϕ10-s10	BamHI (8mer)	ATC	
pET-2b	pAR2106	BglII	ϕ10-s10	BamHI (10mer)	GAT	
pET-2a	pAR2113	BglII	ϕ10-s10	BamHI (12mer)	GGA	
pET-3c	pAR3038	BglII	ϕ10-s10	BamHI (8mer)	ATC	TϕEcoRV

TABLE 5-continued

Vectors for Transcribing and Translating Cloned DNAs

Vector	Isolation No.	Upstream site	Expression signals	Fusion cloning site	Open frame	Down-stream
pET-3b	pAR3039	BglII	ϕ10-s10	BamHI (10mer)	GAT	TϕEcoRV
pET-3a	pAR3040	BglII	ϕ10-s10	BamHI (12mer)	GGA	TϕEcoRV

The parental plasmid for this set of plasmids is pAR2067, which was derived from pBR322 by eliminating the NdeI site at nucleotide 2298 by opening it, filling in, and re-ligating the blunt ends.

A TaqI-RsaI fragment of T7 DNA (nucleotides 22,880–22,998) containing the ϕ10 promoter for T7 RNA polymerase, the translation initiation site for the gene 10 protein has been inserted into the BamHI site of pAR2067. (Gene 10 specifies the major capsid protein of T7.) The fragment extends from nucleotides −23 to +96 relative to the start of the RNA and is oriented in each plasmid so that transcription from the ϕ10 promoter is directed counterclockwise, opposite to transcription from the tetracycline promoter. Initially, the fragment was cloned with a BamHI linker CGGGATCCCG (10 mer) attached to the upstream end and one of three BamHI linkers, CGGATCCG (8 mer), CGGGATCCCG (10 mer), or CGCGGATCCGCG (12 mer), attached to the downstream end of the T7 DNA fragment. This produced a set of three plasmids such that inserting a DNA fragment having an open reading frame into the downstream BamHI site will produce an in-frame fusion to the first 11 amino acids of the gene 10 protein in one of the three plasmids. If insertion is by means of a BamHI-compatible sticky end (GATC), the open reading frame generated by the 8mer linker will be ATC, that by the 10 mer GAT, and that by the 12mer GGA.

To make this set of plasmids convenient to use, the upstream BamHI site was eliminated by opening, filling in, and re-ligating, or was converted to a BglII site by adding the linker GAGATCTC to the filled in site, cutting with BglII and ligating. This leaves the downstream BamHI site as a unique cloning site in these plasmids. New vectors can be created by moving the BglII-BamHI fragment into any unique BamHI site.

A further modification has been to add Tϕ, the transcription terminator for T7 RNA polymerase, just downstream of the BamHI cloning site. A fragment of T7 DNA containing Tϕ (nucleotides 24,106–24,228, where transcription terminates at nucleotide 24,209) was joined to the BamHI cloning site through the sequence [GGATCC]GG-To-CCGGATATCC, where the final C is nucleotide 375 in the BamHI site of pBR322. The downstream linking sequence contains an EcoRV site (GAT'ATC), so the entire [ϕ10-s10-Tϕ] fragment can be removed from this plasmid as a BglII-EcoRV fragment for transfer to other vectors. (These plasmids also contain one other EcoRV site, at nucleotide 187 of pBR322.)

Each of the plasmids in this set contains a unique NdeI site (CA'TATG) that includes the initiating ATG for the gene 10 protein. Any coding sequence that also begins at an NdeI site can be joined to the upstream translation signals for the gene 10 protein, and such a fusion will specify only amino acids of the desired protein. Initiation codons that are not naturally part of an NdeI site could potentially be converted to such by directed mutagenesis.

The second and third codons for the gene 10 protein specify an NheI site (G'CTAGC), which can also be used to link coding sequences to the beginning of the gene 10 protein. (These plasmids also contain one other NheI site, at nucleotide 229 of pBR322.)

EXAMPLE 16

Expression of Various Target DNAs in Vectors for Transcribing or Translating Cloned DNAs

The vectors described in examples 14 and 15 have been used for transcribing and translating many different DNAs in vitro and in *E. coli* cells. Full length poliovirus cDNA was cloned into the EcoRI site of vector plasmid pAR2192 in both orientations and transcribed by purified T7 RNA polymerase to produce full length infectious poliovirus RNA and full length complementary RNA; and the full length cDNA was cloned in the StuI site downstream of the T7 promoter of pAR2369 and transcribed to produce full length, infectious poliovirus RNA having only two additional nucleotides at the 5' end and no more than seven nucleotides past the poly(A) tract at the 3' end, as described in van der Werf, et al., *PNAS, USA* 83, 2330–2334 (1986). DNA specifying the 23S and 5S rRNAs of the rrnB ribosomal operon of *E. coli* was cloned between the XbaI and BamHI sites of plasmid pAR2156 in the correct orientation for synthesis of the rRNAs from the T7 promoter, and DNA specifying the entire rrnB operon except for the natural promoters was cloned in the BamHI site of plasmid pAR2192 in both orientations; complete transcripts were made of each of these cloned DNAs by purified T7 RNA polymerase, and the ribosomal RNAs were produced in vivo in BL21 (DE3) by T7 RNA Polymerase induced by IPTG from chromosomal copy of the T7 RNA polymerase gene under control of the lacUV5 promoter, as described in Steen, et al., *EMBO J.,* 5, 1099–1103 (1986). A fragmet of poliovirus cDNA coding for a protease involved in processing the poliovirus polyprotein was cloned between the BamHI and HindIII sites of plasmid pAR2106 to form an in-frame fusion after the 11th codon of the T7 gene 10 protein; active protease was produced in HMS174 when T7 RNA polymerase was delivered to the cell by infection with CE6, as described in Toyoda, et al., *Cell,* 45, 761–770 (1986). These are only a few examples of the many uses already made of the vectors described in examples 14 and 15 to direct transcription of DNA fragments by purified T7 RNA polymerase or to direct in vivo expression of DNA fragments by T7 RNA polymerase in BL21(DE3) (after induction) or in HMS174 (after infection by CE6).

We claim:

1. An isolated DNA sequence consisting of a cloned T7 RNA polymerase gene encoding an active T7 RNA polymerase.

2. A DNA expression vector encoding an active T7 RNA polymerase.

3. A cell transformed with the DNA expression vector of claim **2**, wherein active T7 RNA polymerase is produced by the cell.

4. The cell of claim **3** which is a prokaryotic cell.

5. The cell of claim **4** which is *E. coli.*

* * * * *

United States Patent [19]

Innis et al.

[11] **Patent Number: 5,075,216**

[45] **Date of Patent: Dec. 24, 1991**

[54] **METHODS FOR DNA SEQUENCING WITH THERMUS AQUATICUS DNA POLYMERASE**

[75] Inventors: **Michael A. Innis**, Moraga; **Kenneth B. Myambo**, Pittsburg; **David H. Gelfand**; **Mary Ann D. Brow**, both of Oakland, all of Calif.

[73] Assignee: **Cetus Corporation**, Emeryville, Calif.

[21] Appl. No.: **249,367**

[22] Filed: **Sep. 23, 1988**

[51] **Int. Cl.**[5] **C12P 19/34**

[52] **U.S. Cl.** **435/6**; 435/91; 435/810; 436/501; 436/808; 536/27; 935/16; 935/17; 935/18; 935/78; 935/88

[58] **Field of Search** 435/6, 91, 810; 436/501, 808; 536/27; 935/16, 17, 18, 78, 88

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,683,202	7/1987	Mullis	435/91
4,795,699	1/1989	Tabor et al.	435/5
4,921,794	5/1990	Tabor et al.	435/91
5,001,050	3/1991	Blanco et al.	435/5

FOREIGN PATENT DOCUMENTS

0258017	3/1988	European Pat. Off. .
351138	1/1990	European Pat. Off. .
9008839	8/1990	PCT Int'l Appl. .

OTHER PUBLICATIONS

Chait et al. (1988) Nature, vol. 333, pp. 477–478.
Simpson et al. (Feb., 1988) Biochem. and Biophys. Res. Comm., vol. 151, No. 1, pp. 487–492.
Miyusawa et al. (1986) Nuc. Acids Res., vol. 14, No. 3, pp. 1319–1324.
Sanger et al., Dec. 1977, *Proc. Natl. Acad. Sci. U.S.A.* 74:5463–5467.
Mills et al., May 1979, *Proc. Natl. Acad. Sci. U.S.A.* 76(5):2232–2235.
Yanisch-Perron et al., 1985, *Gene* 33:103–119.
Barr et al., 1986, *BioTechniques* 4(5):428–432.
Smith et al., Jun. 12, 1986, *Nature* 321:674–679.
Wrischnik et al., 1987, *Nuc. Acids Res.* 15(2):529–542.
Prober et al., Oct. 16, 1987, *Science* 238:336–341.
Ansorge et al., 1987, *Nuc. Acids Res.* 15(11):4593–4602.
Tabor et al., Jul. 1987, *Proc. Natl. Acad. Sci. U.S.A.* 84:4767–4771.
Wong et al., Nov. 26, 1987, *Nature* 330:384–386.
Toneguzzo et al., 1988, *BioTechniques* 6(5):460–469.
Tindall et al., 1988, *Biochem.* 27:6008–6013.
Engelke et al., Jan. 1988, *Proc. Natl. Acad. Sci. U.S.A.* 85:544–548.
Saiki et al., Jan. 29, 1988, *Science* 239:487–491.
Stoflet et al., Jan. 29, 1988, *Science* 239:491–494.
New England BioLab's Catalog Update, Oct. 1987.
Promega advertisement, product update bulletin and Taq DNA polymerase certificate of analysis.
Stratagene advertisement.
Gyllensten et al., Oct. 1988, *Proc. Natl. Acad. Sci. U.S.A.* 85:7652–7656.
Innis et al., Dec. 1988, *Proc. Natl. Acad. Sci. U.S.A.* 85:9436–9440.
Nakamaye et al., preprint.
Hunkapiller, Jun. 2, 1988, *Nature* 333:477–478.
Shengyu et al., May 1987, *Scientia Sinica* XXX(5):503–506.
Heiner et al., May 27, 1988, Preliminary Draft.
McConlogue et al., Oct. 25, 1988, *Nuc. Acids Res.* 16(20):9869.

Primary Examiner—Robert A. Wax
Assistant Examiner—Ardin H. Marschel
Attorney, Agent, or Firm—Stacey R. Sias; Kevin R. Kaster

[57] **ABSTRACT**

Dideoxynucleotide DNA sequencing methods can be dramatically improved by utilizing the DNA polymerase from *Thermus aquaticus* to catalyze the primer extension reactions.

31 Claims, 3 Drawing Sheets

EXHIBIT T

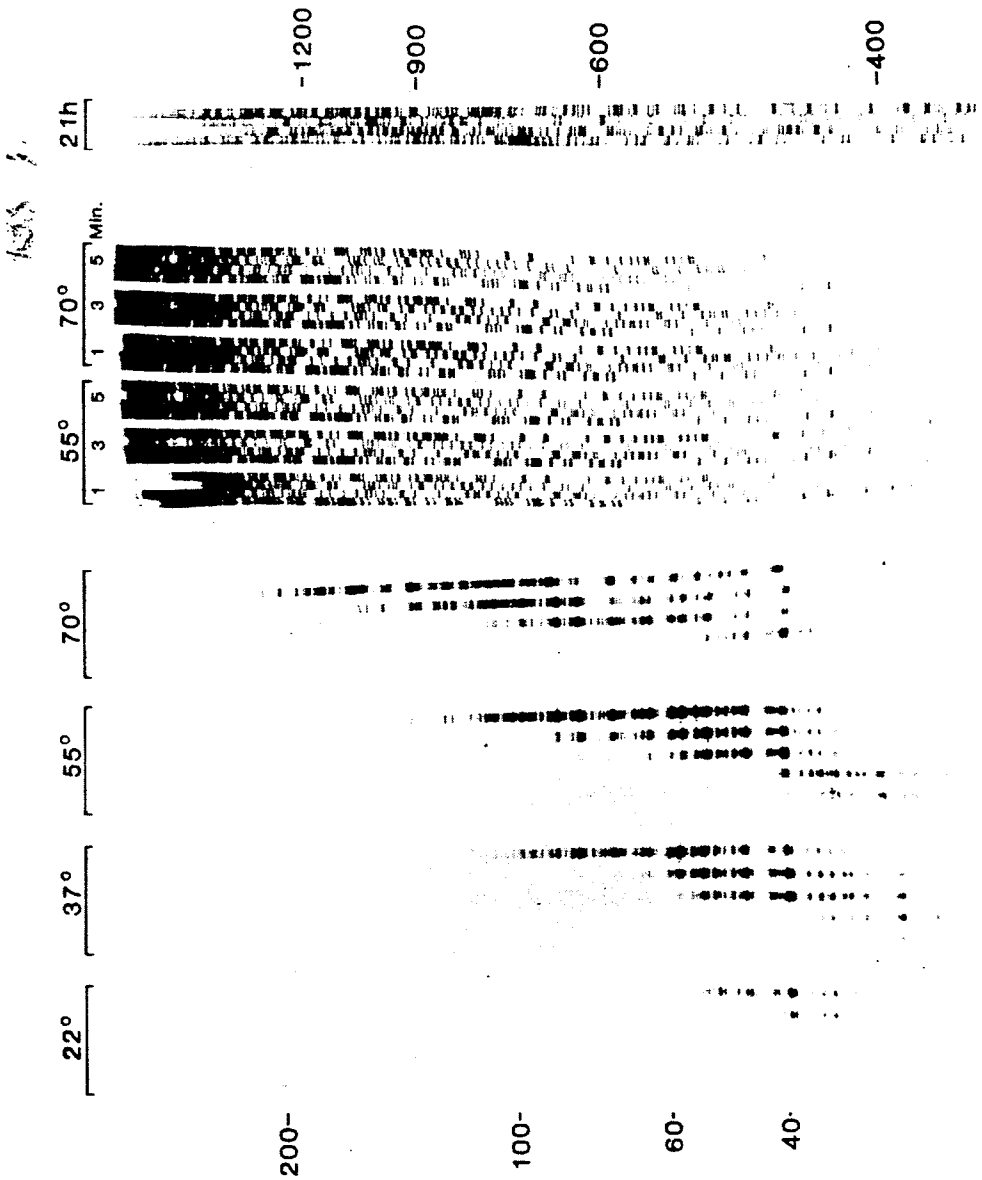


FIG. 1C

FIG. 1B

FIG. 1A

FIG. 2

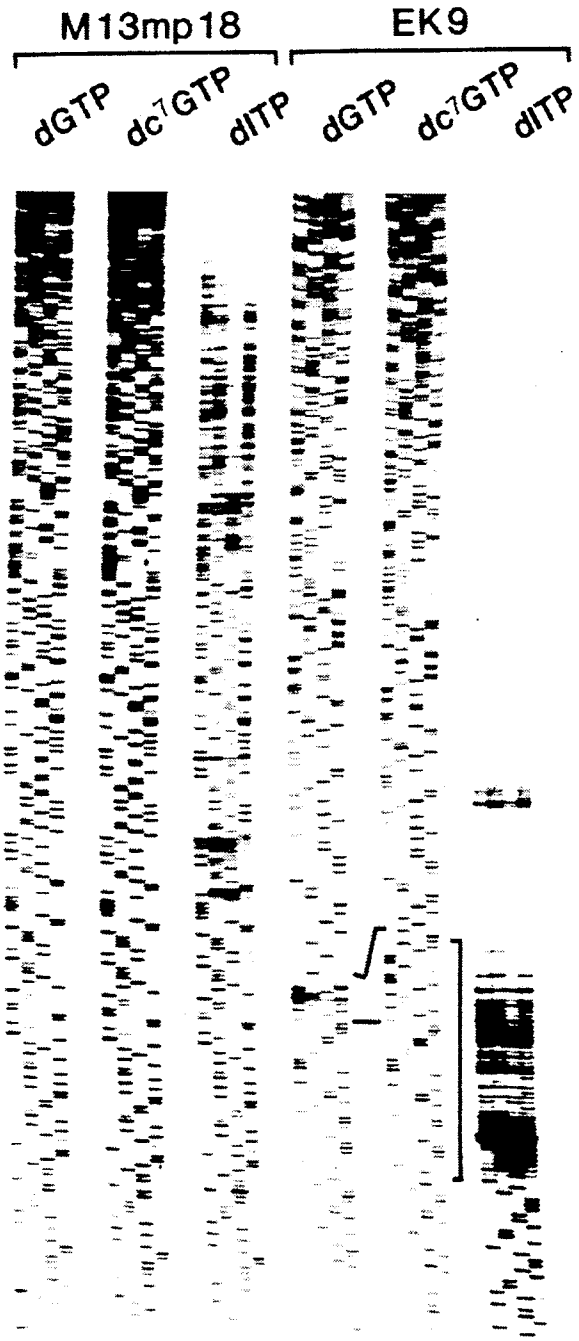


FIG. 3A

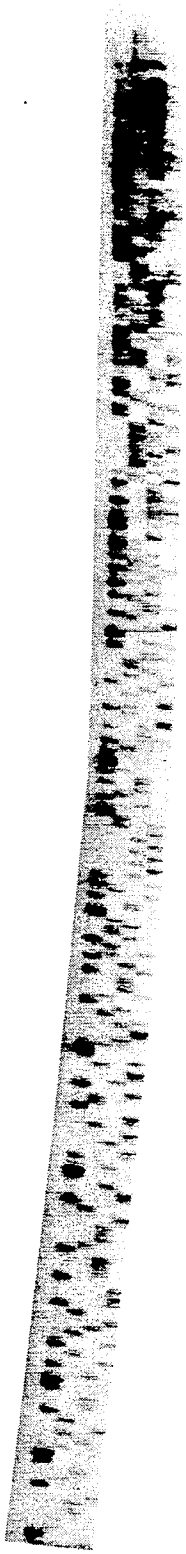
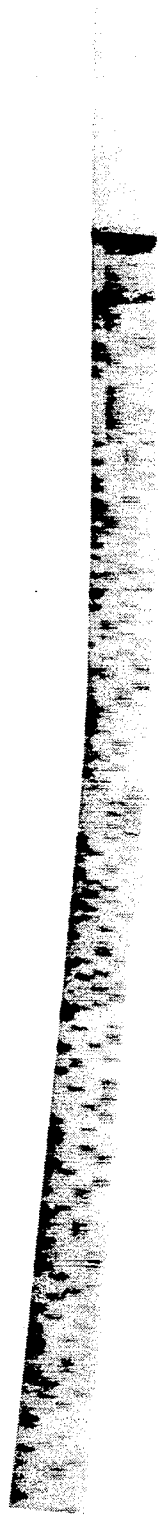


FIG. 3B



METHODS FOR DNA SEQUENCING WITH THERMUS AQUATICUS DNA POLYMERASE

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention provides methods for DNA sequencing utilizing the thermostable DNA polymerase, Taq polymerase, of *Thermus aquaticus*. DNA sequencing methods are of great practical utility in the fields of molecular biology, genetics, medical diagnostic technology, and forensics. The importance of DNA sequencing is evidenced by the significant commercial activity centered about the production and marketing of reagents and automated instruments for sequencing nucleic acids.

2. Description of Related Disclosures

DNA sequencing by the Sanger dideoxynucleotide method (Sanger et al., 1977, *Proc. Natl. Acad. Sci. USA* 74:5463–5467) has undergone significant refinement in recent years, including the development of novel vectors (Yanisch-Perron et al., 1985, *Gene* 33:103–119), base analogs (Mills et al., 1979, *Proc. Natl. Acad. Sci. USA* 76:2232–2235, and Barr et al., 1986, *BioTechniques* 4:428–432), enzymes (Tabor et al., 1987, *Proc. Natl. Acad. Sci. USA* 84:4763–4771), and instruments for partial automation of DNA sequence analysis (Smith et al., 1986, *Nature* 321:674–679; Prober et al., 1987, *Science* 238:336–341; and Ansorge et al., 1987, *Nuc. Acids Res.* 15:4593–4602). The basic dideoxy sequencing procedure involves (i) annealing an oligonucleotide primer to a suitable single or denatured double stranded DNA template; (ii) extending the primer with DNA polymerase in four separate reactions, each containing one α-labeled dNTP or ddNTP (alternatively, a labeled primer can be used), a mixture of unlabeled dNTPs, and one chain-terminating dideoxynucleoside-5'-triphosphate (ddNTP); (iii) resolving the four sets of reaction products on a highresolution polyacrylamide-urea gel; and (iv) producing an autoradiographic image of the gel that can be examined to infer the DNA sequence. Alternatively, fluorescently labeled primers or nucleotides can be used to identify the reaction products. Known dideoxy sequencing methods utilize a DNA polymerase such as the Klenow fragment of *E. coli* DNA polymerase I, reverse transcriptase, or a modified T7 DNA polymerase. Protocols for sequencing with these enzymes, however, do not work with Taq polymerase.

Introduction of commercial kits has vastly simplified the art, making DNA sequencing a routine technique for any laboratory. However, there is still a need in the art for sequencing protocols that work well with nucleic acids that contain secondary structure such as palindromic hairpin loops and with G+C-rich DNA, which can form compressions in the DNA through Hoogsteen bond formation. Such DNA typically performs poorly in prior art sequencing protocols and can exhibit aberrant gel migration patterns that also interfere with sequence determination. In addition, there is a need for sequencing methods that can generate DNA sequence information over a long segment of DNA from one sequencing reaction. Currently, different sequencing methods must be used to generate both short and long sequence products. The present invention, as described more fully below, dramatically improves the art of DNA sequencing by, in one aspect, generating both short and long sequencing products in a single sequencing reaction.

The current commercial instruments address the "backend" of the sequencing process: non-isotopic detection and computerized data collection and analysis. Such developments have led many investigators to undertake large-scale sequencing projects, and to consider the sequencing of the entire human genome. The ultimate success of large-scale sequencing projects will depend upon further improvements in the speed and automation of the technology. These include developing alternative methods for handling the "front-end" of the process, i.e., automating the preparation of DNA templates and the performance of the sequencing reactions, and the present method provides a means for fully automating this frontend of the process.

One technique which appears to be ideally suited for automating DNA preparation is the selective amplification of DNA by the polymerase chain reaction (PCR), a method disclosed in U.S. Pat. No. 4,683,202. Methods for performing PCR are disclosed in pending Ser. No. 063,647, filed June 17, 1987, which is a continuation-in-part (CIP) of Ser. No. 899,513, filed Aug. 22, 1986, now abandoned, which is a CIP of Ser. No. 828,144, filed Feb. 7, 1986, which issued as U.S. Pat. No. 4,683,195, and which is a CIP of Ser. No. 791,308, filed Oct. 25, 1985, which issued as U.S. Pat. No. 4,683,202, and which is a CIP of abandoned Ser. No. 716,975, filed Mar. 28, 1985, all of which are incorporated herein by reference. PCR involves repeated cycles of (i) heat denaturation of the DNA, (ii) annealing of two oligonucleotide primers that flank the DNA segment to be amplified, and (iii) extension of the annealed primers with DNA polymerase. With this method, segments of single-copy genomic DNA can be amplified more than 10 million fold with very high specificity and fidelity. The PCR product can then either be subcloned into a vector suitable for sequence analysis or, alternatively, purified PCR products can be sequenced as disclosed by Engelke et al., 1988, *Proc. Natl. Acad. Sci. USA* 85:544–548; Wong et al., 1987, *Nature* 330:384–386; and Stoflet et al., 1988, *Science* 229:491–494.

Saiki et al., 1988, *Science* 239:487–494, demonstrate that Taq DNA polymerase greatly simplifies the PCR procedure. Because this polymerase has a broad temperature optimum centered around 75° C. and can survive repeated incubations at 95° C., fresh enzyme need not be added after each PCR cycle. Use of Taq DNA polymerase at high annealing and extension temperatures increases the specificity, yield, and length of products that can be amplified, and thus increases the sensitivity of PCR for detecting rare target sequences. Methods for isolating and producing recombinant Taq polymerase are disclosed in pending U.S. patent application Ser. No. 143,441, filed Jan. 12, 1988, which is a CIP of Ser. No. 063,509, filed June 17, 1987, which issued as U.S. Pat. No. 4,889,818, which is a CIP Ser. No. 899,241, now abandoned, filed Aug. 22, 1986, each of which is incorporated herein by reference.

Inverse PCR is a variation of PCR in which the plasmid containing the target template is digested with a restriction endonuclease and recircularized to access flanking sequences for amplification and is fully disclosed in pending Ser. No. 203,000, filed June 6, 1988. PCR has been automated; PCR instruments are disclosed in pending Ser. No. 899,061, filed Aug. 22, 1986, which is a CIP of pending Ser. No. 833,368, filed Feb. 25, 1986, now abandoned. Methods for the structure-

independent amplification of DNA by PCR utilizing the structure-destabilizing base analog 7-deazaguanine are disclosed in pending U.S. Ser. No. 248,556, filed Sept. 23, 1988, and are especially useful in the practice of the present method. Methods for generating single-stranded DNA by a process termed asymmetric PCR are disclosed in pending U.S. Ser. No. 248,896, filed Sept. 23, 1988, and are especially useful in conjunction with the present method. The disclosures of these related patents and applications are incorporated herein by reference.

Prior to the present invention however, Taq DNA polymerase had not been used in DNA sequencing methods. Taq DNA polymerase exhibits high processivity, a rapid rate of incorporation, and ability to utilize nucleotide analogs to terminate chain extension and to resolve gel compressions. These properties of Taq DNA polymerase are similar to those of a chemically modified bacteriophage T7 DNA polymerase recently decribed by Tabor et al., 1987, *Proc. Natl. Acad. Sci. USA* 84:4767–4771. In contrast to T7 DNA polymerase, however, Taq DNA polymerase is a single-chain enzyme which is highly thermostable, as described by Gelfand et al., European Patent Publication 258,017. Because Taq polymerase has no detectible 3'-5'-exonuclease activity, and because the misincorporation rate is high unless certain dNTP and ddNTP concentrations are used, Taq polymerase has not previously been used for sequencing. The present invention provides efficient protocols for DNA sequencing with Taq DNA polymerase, which can also be used for direct sequencing of PCR-amplified DNA.

SUMMARY OF THE INVENTION

The present invention provides an improved dideoxynucleotide method for determining the nucleotide sequence of a nucleic acid. This improved method involves utilization of the DNA polymerase from *Thermus aquaticus*, called Taq polymerase, for the extension of the primers used in the method. The method of the invention is especially preferred when practiced with single stranded DNA generated by a modified or asymmetric polymerase chain reaction to produce single stranded DNA.

The present method offers significant advantages over known sequencing methods. Many of these advantages arise out of special attributes of Taq DNA polymerase, which will not function properly in dideoxy sequencing protocols designed for the Klenow fragment of *E. coli* DNA polymerase I, reverse transcriptase, or a modified form of T7 DNA polymerase. However, using the method of the present invention, one can generate sequence information with Taq polymerase in a manner not possible with any other known protocol.

For instance, the sequencing reaction of the present method can be practiced over a broad range of temperatures, whereas prior art methods were inoperable at temperatures much higher than 50° C. However, at temperatures of 50° C., many single stranded DNAs can form secondary structure, such as a hairpin loop, that can seriously interfere with a dideoxy sequencing protocol, both through improper termination in the extension reaction and through the generation of aberrant migration patterns on sequencing gels. The ability to conduct the extension reaction at a higher temperature, i.e., 70° C., as provided by the present method, results in a significant improvement in sequencing results with DNA that contains such secondary structure, because high temperature destabilizes secondary structure. The ability to use high temperatures with the present invention also results in increased primer specificity, which, in turn, provides cleaner (less background) and more readable sequence information.

The present method also provides better sequencing results because of the ability to utilize structure-destabilizing base analogs such as 7-deazaguanine in the method. This analog can be used to prevent Hoogsteen bond formation in G+C-rich DNA, which, if not prevented, causes compressions in the DNA and aberrant migration patterns of DNA strands on sequencing gels.

Another important advantage of the present method is the ability to generate sequence information over a long segment of nucleotides in a single sequencing reaction (which, as described below, is really 4 different reactions, one for each nucleotide: A, G, C, and T). Taq polymerase is fast and very processive, and products can be generated by the present method that yield signals of uniform intensity, whether the products are short (within 30 nucleotides of the primer) or long (over 1000 nucleotides from the primer). Nor is this advantage limited to sequence determination by the use of autoradiography. Instead, the nature of the generation of extension products in the present method makes possible, for the first time, an automated DNA sequencing instrument capable of determining over 1000 bases of nucleotide sequence in a single sequencing reaction, independent of the method used for detecting extension products. Prior to the present invention, DNA sequencing instruments produced, at best, less than 600 bases of sequence per sequencing reaction.

Another important aspect of the present invention has an even greater impact on DNA sequencing instrumentation. The PCR process has been automated, and asymmetric PCR can be used for generating a single stranded DNA template for sequencing. The Taq polymerase is preferred for PCR, but prior to the present invention, not preferred for DNA sequencing. With the advent of the present invention, however, generation of template for sequencing and determination of sequence can be carried out in a single automated process. The present invention also relates to certain kits and buffers suitable for use in both PCR and the present method.

BRIEF DESCRIPTION OF THE FIGURES

FIG. **1** shows autoradiographs of polyacrylamide-urea gels exhibiting the products of (A) labeling reactions, (B) sequencing (extension-termination) reactions performed at various temperatures, and (C) sequencing reaction products resolved during extended electrophoresis. The labeling reactions were performed as described in Example 4. The reactions were brought up to temperature before the addition of the enzyme. Aliquots were removed at 0.5, 1, 3, 5, 7, and 10 minutes. The extension-termination reactions were performed as described. All reactions were stopped with formamide-EDTA stop solution, denatured at 80° C. for 3 minutes, and resolved on a buffer-gradient sequencing gel (described by Biggin et al., 1983, *Proc. Natl. Acad. Sci. USA* 80:3963–3965). Extended electrophesis (C) was performed on the products of a 70° C./3 minute extension-termination sequencing reaction. Samples were run at 15 W for 21 hours on a 18 cm ×50 cm ×0.4 mm. 7% acrylamide gel (24:1 cross-linking) with 7 M urea and 1X TBE. Markers indicate the distance in nucleotides from the beginning of the primer. All sequencing reaction sets are loaded: G, A, T, C.

FIG. 2 shows an autoradiograph of a polyacrylamide-urea gel comparing extension products generated with base analogs. The effects of replacing dGTP with c^7dGTP (7-deaza-2'-deoxyguanosine-5'-triphosphate is abbreviated in the text as c^7dGTP but is abbreviated in the Figure as dc^{7}GTP) or dITP are shown in sequencing reactions performed on M13:mp18 single-stranded DNA or on a partially palindromic clone, EK9. Lanes are loaded: G, A, T, C. Lines between the EK9 dGTP and c^7dGTP reaction sets align the same positions upstream and downstream of the compressed region. The bracket indicates the limits of the palindrome. The correct sequence of the region is: 5'-CCATGT GACCCTGCCCGACTTCGACG GGAATTCCCGTC"GAAGTCGGGCAGGGT CACC"ATA-3'. The complementary bases are underlined and the bases compressed in the dGTP reactions are in quotes.

FIG. 3 shows an autoradiograph of a polyacrylamide-urea gel on which are compared the extension products from (A) an M13-based single-stranded template, and (B) an asymmetric PCR template of the same sequence. The sequencing of the M13 clone was carried out as described in the accompanying Examples using a [^{32}P]-labeled primer. The asymmetric amplification and subsequent sequencing were performed as described, and all extension products were resolved on a buffer-gradient sequencing gel. Reaction sets were loaded: G, A, T, C.

DETAILED DESCRIPTION OF THE INVENTION

The Sanger and other dideoxynucleoside DNA sequencing protocols involve a series of four reactions, each of which involves the template-dependent extension of an oligonucleotide primer annealed to the nucleic acid to be sequenced, the template. The extension reaction is catalyzed by an agent for template-dependent polymerization. The template DNA is single stranded, so the primer can anneal to the template, and each of the four extension reactions is carried out in the presence of the four dideoxynucleoside-5'-triphosphates (dATP, dCTP, dGTP, and TTP) or in a similar mixture containing one or more natural or synthetic analogs of dATP, dCTP, dGTP, or TTP and one dideoxynucleoside-5'-triphosphate (ddNTP). Incorporation of a ddNTP terminates the extension reaction, and the ddNTP concentration can be adjusted so that the reaction generates molecules of a broad range of chain lengths. The four separate reactions are utilized so that in one reaction all extension products end with ddATP, in another with ddCTP, in another with ddGTP, and in the fourth with ddTTP. Through use of labeled primers, dNTPs, or ddNTPs, the products of the extension reaction can be detected. Separation of the products by size, i.e., on a sequencing gel in side-by-side lanes, and visualization or other detection of the extension reaction products allows the sequence of the template to be determined.

Prior to the present invention, the extension products in dideoxy sequencing methods were generated by agents for polymerization such as the Klenow fragment of *E. coli* DNA polymerase I, reverse transcriptase, or a modified T7 DNA polymerase. The present invention provides a significantly improved method for dideoxynucleotide sequencing that utilizes the DNA polymerase from *Thermus aquaticus*, Taq polymerase, to catalyze the extension reaction.

The present invention provides convenient and efficient methods for sequencing DNA with Taq DNA polymerase. The methods work equally well with either 5'-labeled primers or by incorporation of label in a two-step reaction protocol. Both methods of incorporating label have been used to generate DNA sequencing ladders that are characteristically free of background bands or noticeable enzyme idiosyncrasies, uniform in intensity, and readable over long distances. The present protocols also gave very clean results in sequencing alkali-denatured double stranded DNA templates.

The advantages of the present method will make Taq DNA polymerase the polymerase of choice for most sequencing applications. Sequencing results obtained using the present method were far superior to those obtained using either Klenow or AMV reverse transcriptase methodologies and were better than the results obtained using a method for sequencing with modified T7 DNA polymerase. One reason for these superior results is that, unlike any of these polymerases, Taq DNA polymerase works over a broad temperature optimum centered around 75° C. Regions of DNA secondary structure (hairpins) are commonly encountered and can strongly hinder a DNA polymerase and cause premature termination of the primer extension reaction. This result is observed as bands across all four sequencing lanes on sequencing gels and is cause for failure no matter how extension products are detected. Other structures can interfere in sequencing and are common in high G+C DNA as a result of compression through Hoogsteen bond formation, but can also occur in DNA with no apparent abnormalities. The ability of Taq DNA polymerase to operate at high temperature and low salt allows heat-destablization of hairpins during the sequencing reaction, permitting the enzyme to read through such structures. The concomitant use in the present method of a structure-destabilizing dGTP analog, such as 7-deaza-2'-deoxyguanosine-5'-triphosphate (c^7dGTP), yields sequencing products from such difficult to sequence DNA that were fully resolved upon electrophoresis (see also copending U.S. Ser. No. 248,556, filed Sept. 23, 1988).

Absence of background bands and uniform intensity of the radioactive fragments are benefits provided by the present method. Another benefit is due to the fact that Taq DNA polymerase is very processive. Within two minutes at 70° C., the Taq enzyme can replicate an entire 7.25 kb template. This equals a turnover rate in excess of 60 nucleotides per second. Taq DNA polymerase also has significant activity at lower temperatures with calculated turnover rates of 24, 1.5, and 0.25 nucleotides per second at 55° C., 37° C. and 22° C., respectively. In the absence of ddNTPs, a Taq DNA polymerase extension reaction, at 70° C. and at a substantial substrate excess (0.1:1 molar ratio of polymerase to primer/template), will extend most initiated primers completely prior to reinitiation on new substrate. The extension rate is relatively independent of enzyme concentration and demonstrates that Taq DNA polymerase has high processivity. Taq DNA polymerase also has very little if any proofreading activity.

These properties of Taq enzyme make the present method preferred over other sequencing methods. Polymerase pausing and premature termination of chain extension at sequences with secondary structure is reduced and discrimination against dideoxynucleotide analogs is diminished by the present method. These benefits make the invention preferred for use in auto-

mated sequencing instruments. Ironically, however, one of the beneficial properties of Taq polymerase for sequencing, the absence of significant Taq-associated 3′→5′-exonuclease activity, undoubtedly prevented the development of Taq polymerase sequencing methods even after the purification of the enzyme by Gelfand et al. (European Patent Publication No. EPO 258,017). This is because absence of a 3′→5′-exonuclease activity results in a failure to remove misincorporated bases and results in chain termination. Misincorporation occurs at very low and generally unbalanced nucleotide concentrations typically used in prior art methods. The present inventors discovered that the rate is enhanced unacceptably for sequencing when one or more of the dNTPs are well below Km and/or when the concentration of one dNTP is very low relative to the other dNTPs. The present inventors also discovered that conditions which favor high fidelity and catalytic efficiency over long distances reactions are similar concentrations of each of the four dNTPs and ≧10 μM for each dNTP.

The chain extension reaction conditions of the present sequencing method are especially preferred when carried out in a buffer compatible with PCR, as is discussed more fully below. The buffer (described by Saiki et al., 1988, *Science* 239:487–494) for Taq polymerase PCR reactions (50 mM KCl; 10 mM Tris-HCl, pH 8.4; 2.5 mM $MgCl_2$; 200 μM of each dNTP; and 200 μg/mL of gelatin) was accordingly modified by the present inventors for DNA sequencing. The PCR buffer described by Saiki et al. contains KCl. For purposes of the present method, however, the best extensions occur in the absence of KCl. At 50 mM KCl there was slight inhibition of enzyme activity, and at ≧75 mM KCl, the activity of Taq DNA polymerase was significantly inhibited in the present method. The presence or absence of gelatin, which acts as an enzyme stabilizer in PCR reactions, did not affect the sequencing reactions per se; however, gelatin can cause distortions during electrophoresis. Addition of non-ionic detergents to the enzyme dilution buffer (final concentration of detergent in the sequencing reaction: 0.05% Tween 20 and 0.05% NP40) stimulated the activity of the Taq DNA polymerase and reduced the background caused by false terminations from the enzyme.

Taq DNA polymerase requires free magnesium ion; the concentration of magnesium ion should generally be at least 0.8 mM above that of the dNTPs and ddNTPs present in the sequencing reaction of the present invention. Thus, the preferred PCR buffer for use in conjunction with the present invention does not contain KCl, but does contain 0.05% Tween 20, 0.05% NP40, 3 mM (or higher) $MgCl_2$, in a buffer, 10 mMTris-HCl is preferred, at pH 8.0 to 8.5. The reaction mixtures also contain primer, template, Taq polymerase, dNTPs, and ddNTPs.

The present method can tolerate a wide variety of nucleotide concentrations, especially if each dNTP is present at concentrations greater than 10 μM. However, ddNTPs are expensive, and must be present in the extension reaction at a ratio to the corresponding dNTP for generation of meaningful sequence information. Consequently, lower dNTP concentrations are preferred for any dideoxy sequencing method. At concentrations of less than 5 μM in each of four dNTPs, and when the concentration of one dNTP was low relative to the other dNTPs, a high background of incorrect termination products was seen due to misincorporation of both dNTPs and ddNTPs.

Thus, the optimum concentration for each ddNTP was empirically determined in a solution containing of all four dNTPs, each present at 10 μM. Taq DNA polymerase incorporated the four ddNTPs with varying efficiency, and much less efficiently than the corresponding dNTPs. Ratios that generated optimal distributions of chain termination products were: dGTP:ddGTP (1:6), dATP:ddATP (1:32), TTP:ddTTP (1:48), and dCTP:ddCTP (1:16).

Taq DNA polymerase concentration was varied between 1 and 20 units per set of four reactions containing 0.2 pmol of single stranded DNA template, 0.5 pmol of primer, and the dNTP:ddNTP concentrations described above. The amount of extension products synthesized increased up to 10 units of polymerase per reaction set. At this concentration of reagents, ten units of Taq DNA polymerase represented approximately a 2.5-fold molar excess of enzyme over template-primer; however, a one:one ratio of Taq polymerase: template-primer is less costly and works well.

The present invention also encompasses a variety of methods for incorporating labeled nucleotide during the sequencing reaction. One popular method involves the use of a labeled primer in the sequencing (chain extension and termination) reactions. Another method involves incorporation of a labeled nucleotide into the extending primer. A Klenow-type protocol, where one labeled nucleotide is present at low concentration relative to the other three during primer extension, however, will not work with Taq polymerase due to misincorporation of dNTPs and ddNTPs. The apparent Km values for each of the four dNTPs is between 10 μM and 20 μM. When the concentration of one labeled nucleotide, either {α-[^{35}S]thio}dATP, or {α-[^{35}S]thio}dCTP, was significantly below Km (i.e., about 0.5 to 1 μM), ddNTPs present at 80–500 μM were inappropriately incorporated at high frequency with Taq Polymerase. Concentrations higher than 1 μM for an [α-^{35}S]-labeled dNTP are not practical. Also, because the Taq enzyme apparently lacks 3′→5′-exonuclease (proofreading) activity, misincorporated dNTPs induce chain termination.

To circumvent these problems and realize the full benefits of dideoxy sequencing with Taq polymerase, the present invention provides a two-step procedure involving an initial low temperature labeling step using uniformly low concentrations of all four dNTPs (one of which is labeled) followed by the sequencing reaction step in the presence of ddNTPs and higher dNTP concentrations. The sequencing reaction can be performed at higher temperatures to achieve superior results. To obtain sequence data in the region next to the primer utilizing this labeling procedure, it is preferred to use both low temperature and limiting dNTP concentrations to generate an array of radioactive extension products ranging in size from a few to greater than 100 nucleotides in length. Minimum concentrations of 0.5 μM for each labeled dNTP are preferred in this step to generate easily readable signals from an overnight exposure, and increasing the concentration of one unlabeled dNTP to 1.0 μM make the signals very clear. This benefit is seen regardless of which dNTP is increased, but increasing more than one is not necessary.

After the labeled nucleotide is incorporated to readable levels, the sequencing reaction is initiated by the addition of balanced dNTPs (≧10 μM each) and

ddNTPs. During the sequencing reactions, temperature increase and higher dNTP concentrations ensure maximum processivity and fidelity with the present method. Sequencing reactions work well in a broad temperature range. The reactions performed at 55° C. occurred at a slower rate, consistent with the extension rates described above, but there was no detectable difference in fidelity as compared with 70° C. Under these conditions, there was remarkable uniformity in the band intensities, and no detectable idiosyncratic band patterns. In addition, the same reaction conditions cover both short and long gel runs. DNA sequence information in excess of 1000 nucleotides from the priming site can be generated using the present method, as shown in FIG. 1(*c*).

The present method can also be carried out using the base analog 7-deaza-2'-deoxyguanosine-5'-triphosphate (c^7dGTP) and high temperature to sequence through G+C-rich DNA and to eliminate band compressions. Band compressions resulting from abnormal gel migration of certain sequences are frequently encountered with G+C-rich DNA templates and occur even in cloned DNA sequences with no apparent abnormality in base composition. Such compressions can result in inconclusive or error-prone reading of sequencing gels. Substitution of dGTP with dITP or c^7dGTP, has been somewhat useful in resolving compression artifacts in known sequencing protocols. Incorporation of such nucleotide triphosphate analogs by Taq polymerase in the present method was investigated using either an M13:mp18 template or a G+C-rich, strong dyad symmetry-containing insert cloned into M13, as shown in FIG. 2. Taq DNA polymerase incorporated c^7dGTP with essentially the same kinetics as dGTP, and a combination of high reaction temperature and c^7dGTP is very efficient for resolving difficult sequences.

In contrast, inosine-containing reactions required a 4-fold higher level of dITP as compared to dGTP, the labeling reaction needed 4 minutes, and the ratio of ddGTP to dITP was reduced 20-fold compared to dGTP. Because deoxyinosine-5-triphosphate (dITP) base-pairs promiscuously, frequent chain termination at regions of secondary structure occur with dITP, which is therefore not preferred for purposes of the present invention. Terminations caused by inosine result both from a higher rate of misincorporation with dITP as compared to the other dNTPs and from the fact that Taq DNA polymerase lacks the necessary 3'→5'-exonuclease activity for editing misincorporated bases. Terminations induced by dITP are greatly reduced if the reactions are initiated at 70° C.

Development of a procedure for direct sequencing of PCR products has been needed since the inception of the PCR technique. The remarkable DNA sequencing results obtained by the present invention, coupled with the compatibility of the present method with PCR, makes the present method the ideal method for directly analyzing PCR products (see FIG. 3). Sequence analysis of cloned PCR products by the present method suggests that the fidelity for PCR using 50–200 μM of each dNTP is quite respectable (approximately one mistake in 4000 nucleotides sequenced following 35 cycles of PCR and cloning of the PCR products) and is comparable with that observed using other DNA polymerases for PCR. Additionally, most misincorporation errors that may occur in the PCR reaction will cause chaintermination, thus preventing amplification of defective molecules.

The present method is especially preferred for use with asymmetric PCR reactions using primers at concentrations designed to first amplify and then generate single stranded DNA (ssDNA) from any insert. Generation of single stranded DNA by a process termed asymmetric PCR is described in pending U.S. patent application Ser. No. 248,896, filed Sept. 23, 1988. This embodiment of the invention was illustrated by cloning DNA into the M13/pUC-lacZ polylinker and generating ssDNA by asymmetric PCR. Asymmetric PCR was performed, as described in the accompanying examples, with one of the oligonucleotide primers present in a 100-fold greater concentration than the other, so that one of the two PCR primers was depleted during the early thermal cycles. The reaction generated single stranded product from the remaining primer.

Sequencing of asymmetric PCR-generated templates by the present method did not require purification of the product. Based on an estimated yield of 1 μg of single stranded product, one-third to one-half of the 2 nmol of each dNTP initially added are used up during the PCR cycles. In addition, the stability of the dNTPs during PCR was determined to be approximately 50% after 60 cycles of PCR. Accordingly, the termination mixes used in the present method are formulated to boost the dNTPs to a final concentration of about 10 μM or higher in the sequencing reaction, to supply specific ddNTPs at appropriate concentrations as determined above, and to provide additional DNA polymerase. A [^{32}P]-labeled sequencing primer can be used to avoid purifying the PCR product and to simplify the sequencing protocol to a single extension/termination step. It is obvious that a fluorescently-labeled sequencing primer(s) could also be used in the present methods; the products can then be analyzed on an automated DNA sequencing instrument.

DNA sequence obtained with Taq DNA polymerase using either an asymmetric PCR-generated template or the same DNA insert cloned in M13:mp18 as template was compared. The resulting sequence ladders showed the clarity and uniformity of signal characteristic of Taq-generated sequences. Any degradation of enzyme or dNTPs that may have occurred during the PCR thermal cycling did not seem to affect the generation of clean sequence data. Synthesis of single-stranded DNA template during 35 cycles of PCR was largely independent of the initial DNA concentration. Asymmetric PCR reactions performed using either 0.1 to 100 ng of M13:mp10 ssDNA, or 10 μL of an M13 phage plaque picked directly into 100 μl of water, sequenced equivalently using the method of the invention.

Although the present invention is illustrated below by sequencing inserts cloned into M13/pUC-based vectors, the method is applicable to direct sequencing of clones in lambda phage or any other cloning vector. Some variability in the ssDNA yield of the asymmetric PCR reaction has been observed for different primer pairs and ratios, and the reaction conditions for each amplification system will need to be adjusted to give the optimum results for a particular primer pair and template nucleic acid. The PCR dNTP concentrations may also need to be varied for products of different sizes and/or amplification efficiencies. Additionally, some investigators have increased the homogeneity of PCR products from genomic DNA by electrophoretic separation and reamplification of eluate from a selected gel slice. The present sequencing method is easily applied to this "secondary" form of PCR. Direct sequencing of

PCR products from DNA by any method produces a "consensus" sequence; those bases which occur at a given position in the majority of the molecules will be the most visible on an autoradiograph and any low-frequency errors will be undetectable. In such a PCR-coupled embodiment of the present invention, the resulting sequence data will be only as clean as the amplified product. Heterogeneous products will naturally produce mixed ladders.

Because Taq DNA polymerase is very useful for PCR, the present invention makes possible the coupling of template preparation by PCR with direct sequencing. This advantage is significant in that it is now possible, by virtue of the present method, to automate both DNA template preparation by PCR and performance of the sequencing reactions in a manner compatible with current partially automated DNA sequencing instruments.

Those skilled in the art recognize that the present method can be used in a variety of contexts where determination of DNA sequence information is desired. The following examples are provided merely to illustrate the invention and not to limit the scope of the accompanying claims. Example 4 presents a preferred embodiment of the present invention.

EXAMPLE 1

Annealing, Labeling, and Extension-Termination Reactions

The materials used in the procedures described below were obtained as follows. Polynucleotide kinase from T4-infected *E. coli* cells was purchased from Pharmacia. Taq DNA polymerase, a single subunit enzyme, was purified from *Thermus aquaticus*, strain YT-1 (ATCC #2543). More recently, Taq DNA polymerase was purchased from Perkin Elmer-Cetus Instruments. The polymerase (5–80 units/μL) was stored at −20° C. in 20 mM Tris-HCl, pH 8.0; 100 mM KCl; 0.1 mM EDTA; 1 mM DTT; 200 μg/mL autoclaved gelatin; 0.5% NP40; 0.5% Tween-20; and 50% glycerol. The enzyme has an approximate specific activity of 200,000 units/mg, with one unit corresponding to 10 nmol of product synthesized in 30 minutes using activated salmon sperm DNA. 2′-deoxy and 2′,3′-dideoxynucleotide-5′-triphosphates (dNTPs and ddNTPs) were obtained from Pharmacia. 7-deaza-2′-deoxyguanosine-5′-triphosphate (c^7dGTP) was from Boehringer Mannheim. {α-[^{35}S]thio}dATP (650 Ci/mmol) was from Amersham, and γ-[^{32}P]ATP was from New England Nuclear. Oligonucleotide primers for sequencing were synthesized on a Biosearch 8700 DNA Synthesizer using cyanoethyl phosphoramidite chemistry. Oligonucleotide primers were 5′-end labeled (3×10^6 cpm/pmol) with γ-[^{32}P]ATP and T4 polynucleotide kinase (Maxam and Gilbert, 1980, *Methods Enz.*, 65:499–560). Single stranded M13 DNA templates were prepared as described by Zinder et al., 1982, *Gene*, 19:1–10.

Single annealing and labeling reactions were performed for each set of four sequencing reactions in 1.5 mL microfuge tubes. The annealing mixture contained 5 μL of oligonucleotide primer (0.1 pmol/μL) in 6 x Taq Sequencing Buffer (TSB, 10 mM $MgCl_2$ and 10 mM Tris-HCl, pH 8.0, at room temperature) and 5 μL of template DNA (0.05 to 0.5 pmol). The mixture was heated in a boiling water bath for three minutes, incubated at 42° C. for 20 minutes, cooled to room temperature, and briefly spun to collect the fluid at the bottom of the tube.

To the 10 μL annealing reaction were added 2 μL of labeling mix (10 μM dGTP, 5 μM dCTP, and 5 μM TTP in 10 mM Tris-HCl, pH 8.0), 2 μL of {α-[^{35}S]thio}dATP (5 μM after 3× dilution in 10 mM Tris-HCl, pH 8.0), 2 μL of Taq DNA polymerase (5 units/μL in dilution buffer: 10 mM Tris-HCl, pH 8.0; 0.5% Tween 20; and 0.5% NP40), and 4 μL H_2O. The labeling reaction was incubated for one minute at 37° C. For sequencing with 5′-labeled primers, the addition of {α-[^{35}S]thio}dNTP, labeling mix, and the labeling reaction were omitted, and the volume was made up with 10 mM Tris-HCl, pH 8.0.

Four separate sequencing (extension-termination) reactions were performed in 96-well microtiter plates (Falcon #3911) for each labeled template using concentrated deoxy/dideoxy termination mixes as follows: "G-mix" (30 μM in each dNTP, 0.25 mM ddGTP, and 0.37 mM $MgCl_2$); "A-mix" (30 μM in each dNTP, 1.0 mM ddATP, and 1.12 mM $MgCl_2$); "T-mix" (30 μM in each dNTP, 1.5 mM ddTTP, and 1.62 mM $MgCl_2$); and "C-mix" (30 μM in each dNTP, 0.5 mM ddCTP, and 0.62 mM $MgCl^2$). Four μL aliquots from the labeling reactions were added at room temperature to wells containing 2 μL of the appropriate termination mix. Reactions were overlaid with 10 μL of mineral oil to prevent evaporation and then incubated at 70° C. for one to three minutes. Reactions were stopped by the addition of 2 μL of 95% deionized formamide containing 0.1% bromophenol blue, 0.1% xylene cyanol, and 10 mM EDTA, pH 7.0. Samples were heated at 80° C. for three minutes before loading 1 to 2 μL onto a buffer-gradient sequencing gel, as described by Biggin et al., *Proc. Natl. Acad. Sci. USA* 80:3963–3965. Results are shown in FIG. **1**.

EXAMPLE 2

Asymmetric Polymerase Chain Reactions

This example describes how DNA can be generated for sequencing by the present method. The template for asymmetric PCR reactions was single stranded M13:mp10 DNA containing a 400 base insert in the EcoRI site of the polylinker. Oligonucleotides (20-mers) were synthesized to flank the polylinker immediately outside of the universal "-20" and "Reverse" sequencing primer binding sites, and these primers were designated RG05 (5′AGGGTTTTCCCAGTCACGAC3′) and RG02 (5′GTGTGGAATTGTGAGCGGAT3′), respectively. Each PCR reaction contained 20 pmol of one primer and 0.2 pmol of the other primer, 20 μM of each dNTP, 1 to 10 ng of DNA, 1X modified PCR buffer (10 mM Tris-HCl, pH 8.0; 3.0 mM $MgCl_2$, and 0.05% of each of Tween 20 and NP40), and 2.5 units of Taq DNA polymerase in a total volume of 100 μL. The reactions were overlaid with 75 μL of mineral oil to prevent evaporation.

Reactions were performed in 0.5 mL microcentrifuge tubes using the Perkin Elmer-Cetus Thermal Cycler. The programmed thermal profile was initiated with a denaturation at 93° C. for 30 seconds, cooled for primer annealing at 50° C. for one minute, heated up to the 72° C. extension temperature over the course of 1.5 minutes, and held at 72° C. for one minute to ensure completed extension. This profile was repeated for 35 cycles, and the final 72° C. incubation was extended to 10 minutes.

EXAMPLE 3

Sequencing of PCR Products

Aliquots of PCR reactions were directly incorporated into dideoxy chain-termination sequencing reactions. A set of four, base-specific chain-termination sequencing mixes was made up, each in 1X modified PCR Buffer and 20 μM of each dNTP. The individual mixes contained 250 μM ddGTP, 1.28 mM ddATP, 1.92 mM ddTTP, or 640 μM ddCTP. For each PCR product to be sequenced, four wells on a 96-well microtiter plate were labeled "G", "A", "T", or "C", and each well received 2.5 μL of the appropriate sequencing termination mix. A 20 μL aliquot of each PCR reaction was removed to a 1.5 mL microcentrifuge tube and mixed with 0.5 μL of fresh Taq DNA polymerase (48 units/μL), 1 μL of the appropriate [^{32}P]-labeled M13 forward or reverse sequencing primer (5'GTAAAACGACGGCCAGT3' and 5'AACAGCTATGACCATG3', respectively, 1.2 pmol per μL) and 10.5 μL of 1X modified PCR buffer. The PCR/primer preparation was immediately distributed in 7.5 μL aliquots into the wells containing the termination mixes and mixed with the pipettor. Samples were overlaid with 10 μL of mineral oil, and the plate was spun to collect the reaction mixture and to distribute the oil across the wells in an even layer. The reactions were incubated at 70° C. for two minutes and stopped by the addition of 4 μL of 91% formamide with 20 mM EDTA pH 8.0, and 0.05% each of xylene cyanol and bromophenol blue. 5 μL aliquots of these reactions were heated to 75° C. for five minutes, and 1 to 2 μL were loaded on a buffer gradient sequencing gel. The results are shown in FIG. 3.

EXAMPLE 4

Preferred Sequencing Protocol

A. ANNEALING TEMPLATE AND PRIMER

Combine in a 1.5 ml microcentrifuge tube: 5 μL of template DNA (0.5 pmol); 1 μL of primer (0.5 pmol); and 4 μL of 5X Sequencing Buffer. The total volume should be 10 μL; if a smaller volume of DNA is used, make up the difference with distilled water. Heat the tube at 70° C. for 3 minutes, then at 42° C. for 10 minutes.

B. LABELING REACTION

Dilute the Taq DNA polymerase enzyme 1:10 in Enzyme Dilution Buffer to 5 U/μL; keep on ice. To the annealed template/primer add the following: 2 μL of Labeling Mix (dGTP or c^{7}dGTP); 1 μL of {α-[^{35}S]thio}dATP (>600 Ci/mmol; diluted to 10 μM in 10 mM Tris-HCl, pH 8.5); 5 μL of distilled water; and 2 μL of diluted Taq DNA polymerase (5 U/μL). Vortex briefly to mix, collect by spinning the tube in a microfuge, and incubate at 37° C. for 2 minutes.

Note that the c^{7}dGTP Labeling Mix should be used if c^{7}dGTP will be used in the sequencing reactions. Use of c^{7}dGTP is recommended for resolving sequences which cause compressions on the gel. Termination Mixes should be aliquoted into the microtiter plate wells prior to starting the labeling reaction. If labeled primers are to be used for sequencing, the {α-[^{35}S]thio}dATP, Labeling Mix, and the labeling reaction incubation are omitted, and the volume is brought to 20 μl with 10 mM Tris-HCl, pH 8.5.

C. TERMINATION REACTIONS

The sequencing termination reactions may be carried out in a microtiter plate (Falcon #3911), using 4 wells per template/primer, labeled G, A, T, and C. Place 2 μL of the ddGTP Termination Mix in the well labeled G. Similarly, place 2 μL of the ddATP, ddTTP, and ddCTP Termination Mixes in the appropriately labeled wells. Note that the c^{7}dGTP Termination Mixes should be used if c^{7}dGTP was used in the labeling reaction.

Immediately upon completion of the labeling reaction, transfer 4 μL aliquots to each of the four wells labeled G, A, T, and C. Place the drops on the sides of the wells, and allow them to slide down to mix with the Termination Mixes. When all wells for all reactions are filled, briefly spin the microtiter plate to ensure that mixing of the labeling reactions with the Termination Mixes is complete.

Incubate the microtiter plate, at 70° C. for 2 minutes, using a heat block which will contact the well bottoms. This time is sufficient to generate extension products greater than 1500 bp in length. Longer times lead to excessive evaporation.

D. STOPPING REACTIONS

Remove the reactions from the heat block and place at room temperature. Add 2 μL of Stop Solution to the side of each well. Briefly spin the microtiter plate to mix the Stop Solution with the reactions. Samples may be stored covered at −20° C. for up to 7 days with minimal degradation.

Immediately prior to loading samples on the gel, heat to 70° C. for 4 minutes. Load 1 to 2 μL per sample in each lane.

E. REAGENTS

Taq DNA Polymerase Sequencing Buffer (5× Concentrate) is 50 mM Tris-HCl, pH 8.5; and 30 mM $MgCl_2$.

Enzyme Dilution Buffer is 10 mM Tris-HCl, pH 8.0; 0.5% Tween 20; and 0.5% NP40.

Labeling Mix (c^{7}dGTP) is 10 μM c^{7}dGTP; 5 μM dCTP; and 5 μM TTP.

Labeling Mix (dGTP) is 10 μM dGTP; 5 μM dCTP; and 5 μM TTP.

ddG Termination Mix (for c^{7}dGTP) is 60 μM c^{7}dGTP; 30 μM in each of dATP, TTP, and dCTP; and 180 μM ddGTP.

ddG Termination Mix (for dGTP) is 30 μM in each dNTP and 180 μM ddGTP.

ddA Termination Mix (for c^{7}dGTP) is 60 μM c^{7}dGTP; 30 μM in each of dATP, TTP, dCTP; and 1 mM ddATP.

ddA Termination Mix (for dGTP) is 30 μM in each dNTP and 1 mM ddATP.

ddT Termination Mix (for c^{7}dGTP) is 60 μM c^{7}dGTP; 30 μM in each of dATP, TTP, and dCTP; and 1.5 mM ddTTP.

ddT Termination Mix (for dGTP) is 30 μM in each dNTP and 1.5 mM ddTTP.

ddC Termination Mix (for c^{7}dGTP) is 60 μM c^{7}dGTP; 30 μM in each of dATP, TTP, and dCTP; and 500 μM ddCTP.

ddC Termination Mix (for dGTP) is 30 μM in each dNTP and 500 μM ddCTP.

Taq DNA polymerase enzyme is stored at a concentration of 50 U/μL.

Stop Solution is 95% formamide; 20 μM ESTA; 0.1% bromophenol blue; and 0.1% xylene cyanol.

Other modifications of the embodiments of the invention described above that are obvious to those of ordinary skill in the areas of molecular biology, medical diagnostic technology, biochemistry, and related disci-

plines are intended to be within the scope of the accompanying claims.

We claim:

1. In a method for determining a nucleotide sequence for a nucleic acid segment by a dideoxynucleoside-5'-triphosphate chain termination procedure, wherein said sequence is determined by extending an oligonucleotide primer in a templatedependent manner in the presence of an agent for polymerization, four dideoxynucleoside- 5'-triphosphates (dNTPs), and a dideoxynucleoside- 5'-triphosphate (ddNTP), the improvement comprising extending said primer in the presence of the agent for polymerization that is *Thermus aquaticus* DNA polymerase.

2. The method of claim 1, wherein said primer is labeled.

3. The method of claim 1, wherein one of the four dNTPs or ddNTP is labeled.

4. The method of claim 1, wherein the four dideoxynucleoside-5'-triphosphates are dATP, dCTP, dGTP, and TTP.

5. The method of claim 1, wherein the four dideoxynucleoside- 5'-triphosphates are dATP, dCTP, c^7dGTP, and TTP.

6. The method of claim 1, wherein the four dideoxynucleoside- 5'-triphosphates are dATP, dCTP, dITP, and TTP.

7. The method of claim 1, wherein said nucleic acid segment was produced by an asymmertric polymerase chain reaction.

8. The method of claim 1, wherein said nucleic acid segment was produced by an asymmetric polymerase chain reaction.

9. The method of claim 1, wherein no KCl is present in the reaction mixture.

10. The method of claim 1, wherein the DNA polymerase is present in up to a 2.5 -fold molar excess over the nucleic acid segment.

11. The method of claim 2, wherein said primer is labeled with ^{32}P, ^{35}S, or a fluorescent molecule.

12. The method of claim 3, wherein said dNTP or ddNTP is labeled with ^{35}S, ^{32}P, or a fluorescent molecule.

13. The method of claim 3, wherein said extension reaction is carried out first at a low temperature and in the presence of three unlabeled dNTPs and one labeled dNTP each present at a concentration of less than 1 μM and then at higher temperatures in higher concentrations of the unlabeled dNTPs.

14. The method of claim 4, wherein said dideoxynucleoside- 5'-triphosphate is ddATP.

15. The method of claim 4, wherein said dideoxynucleoside- 5'-triphosphate is ddCTP.

16. The method of claim 4, wherein said dideoxynucleoside- 5'-triphosphate is ddGTP.

17. The method of claim 4, wherein said dideoxynucleoside- 5'-triphosphate is ddTTP.

18. The method of claim 4, wherein c^7dGTP is also present during primer extension.

19. The method of claim 4, wherein each dNTP is present at a concentration of 5 μM to 30 μM.

20. The method of claim 13, wherein the concentration of said labeled dNTP is 0.5 μM and the concentration of each unlabeled dNTP is 1.0 μM during said low temperature extension reaction.

21. The method of claim 14, wherein the dATP:-ddATP ratio is 1:32.

22. The method of claim 15, wherein the dCTP:ddCTP ratio is 1:16.

23. The method of claim 16, wherein the dGTP:ddGTP ratio is 1:6.

24. The method of claim 17, wherein the TTP:ddTTP ratio is 1:48.

25. The method of claim 19, wherein the concentration of each dNTP is 10 μM.

26. A kit for determining a nucleotide sequence for a nucleic acid segment by a dideoxynucleoside-5'-triphosphate chain termination procedure, which kit comprises

(a) a primer for extending, in a template-dependent manner, a nucleic acid comprising a sequence complementary to said nucleic acid segment;

(b) four deoxyribonucleoside-5'-triphosphates (dNTPs);

(c) four dideoxyribonucleoside-5'-triphosphates (ddNTPs); and

(d) Taq polymerase.

27. The kit of claim 26, wherein said dNTPs are dGTP, dATP, TTP, and dCTP, and said ddNTPs are ddGTP, ddATP, ddTTP, and ddCTP.

28. The kit of claim 26 that comprises c^7dGTP.

29. The kit of claim 26, wherein said dNTPs are described as a set of four dNTPs selected from the group consisting of:

dGTP, dATP, dCTP, and TTP;

c^7dGTP, dATP, dCTP, and TTP; and

dITP, dATP, dCTP, and TTP.

30. The kit of claim 26, wherein said primer hybridizes to an M13 cloning vector and is suitable for sequencing cloned inserts in M13 and/or pUC-based vectors.

31. The kit of claim 27, that comprises:

(i) a G-termination mix comprising dGTP, dATP, TTP, dCTP, and ddGTP;

(ii) a A-termination mix comprising dGTP, dATP, TTP, dCTP, and ddATP;

(iii) a T-termination mix comprising dGTP, dATP, TTP, dCTP, and ddTTP; and

(iv) a C-termination mix comprising dGTP, dATP, TTP, dCTP, and ddCTP.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 5,075,216

DATED : December 24, 1991

INVENTOR(S) : Innis, et al

It is certified that error appears in the above-indentified patent and that said Letters Patent is hereby corrected as shown below:

In column 5, in line 42, delete "dideoxynucleoside" and insert therefor --deoxynucleoside--.

In column 14, in line 63, delete "ESTA" and insert therefor --EDTA--.

In column 15, in line 8, delete "templatedependent" and insert therefor --template-dependent--;
in line 9, delete "dideoxynu-" and insert therefor --deoxynu--;
in line 19, delete "dideox" and insert therefor --deox--;
in line 22, delete "dideox" and insert therefor --deox--;
in line 25, delete "dideox" and insert therefor --deox--;
in line 29, delete "an asymmertric" and insert therefor --a--.

Signed and Sealed this

First Day of November, 1994

Attest:

BRUCE LEHMAN

Attesting Officer

Commissioner of Patents and Trademarks

United States Patent [19]

Dockrill et al.

[11] **Patent Number: 5,035,936**

[45] **Date of Patent: Jul. 30, 1991**

[54] **LOOSE FILL INSULATION PRODUCT COMPRISING MINERAL WOOL NODULES**

[75] Inventors: **Mark B. Dockrill**, Clearwater; **John Buckham**, Byron; **Anthony P. Shen**, Sarnia, all of Canada

[73] Assignee: **Fiberglas Canada Inc.**, Point Edward, Canada

[21] Appl. No.: **551,241**

[22] Filed: **Jul. 10, 1990**

Related U.S. Application Data

[62] Division of Ser. No. 307,495, Feb. 8, 1989.

[51] **Int. Cl.**[5] **E04B 1/62;** F16L 59/00; F16L 59/04

[52] **U.S. Cl.** **428/96;** 156/71; 428/360; 428/361

[58] **Field of Search** 427/220; 428/96, 360, 428/361

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,233,433	3/1941	Smith	427/215
3,014,872	7/1975	Scott	252/62
3,894,314	7/1975	Nayba	427/212
4,418,103	11/1983	Toni et al.	428/360
4,542,044	9/1985	Gano et al.	427/215
4,820,574	4/1989	Tesch	428/362

Primary Examiner—James C. Cannon

Attorney, Agent, or Firm—Dennison, Meserole, Pollack & Scheiner

[57] **ABSTRACT**

A nodulizing machine for use in an apparatus for making loose fill insulation from mineral fibers comprises an input portal, a conduit containing a plurality of blades and an output portal. Apparatus for making loose fill insulation from mineral fibers comprises a duct extending from a source of the mineral fibers to the nodulizing machine of the present invention. The apparatus includes flow rate control means to control the flow rate of air and mineral fibers in the apparatus. A process for making loose fill insulation from mineral fibers in association with the apparatus of the invention is also discussed.

4 Claims, 2 Drawing Sheets

EXHIBIT U

Fig. 1.

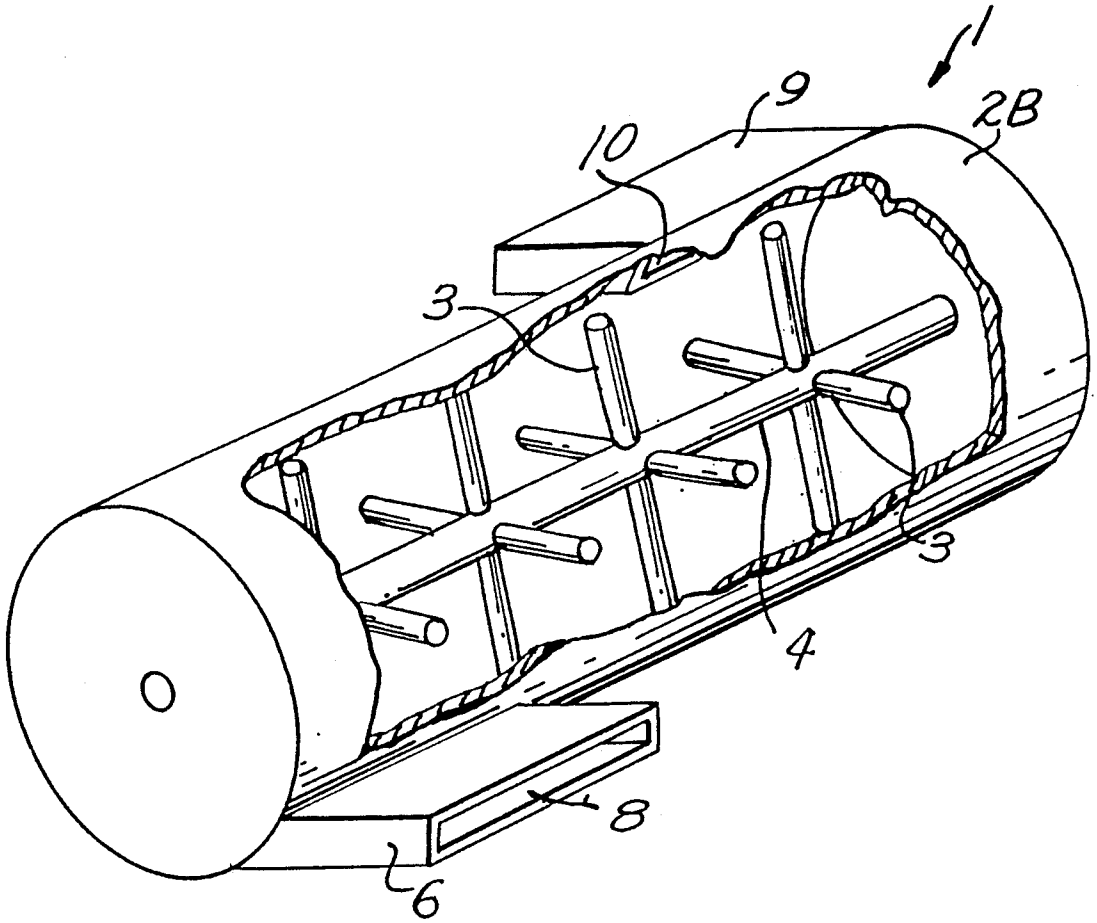
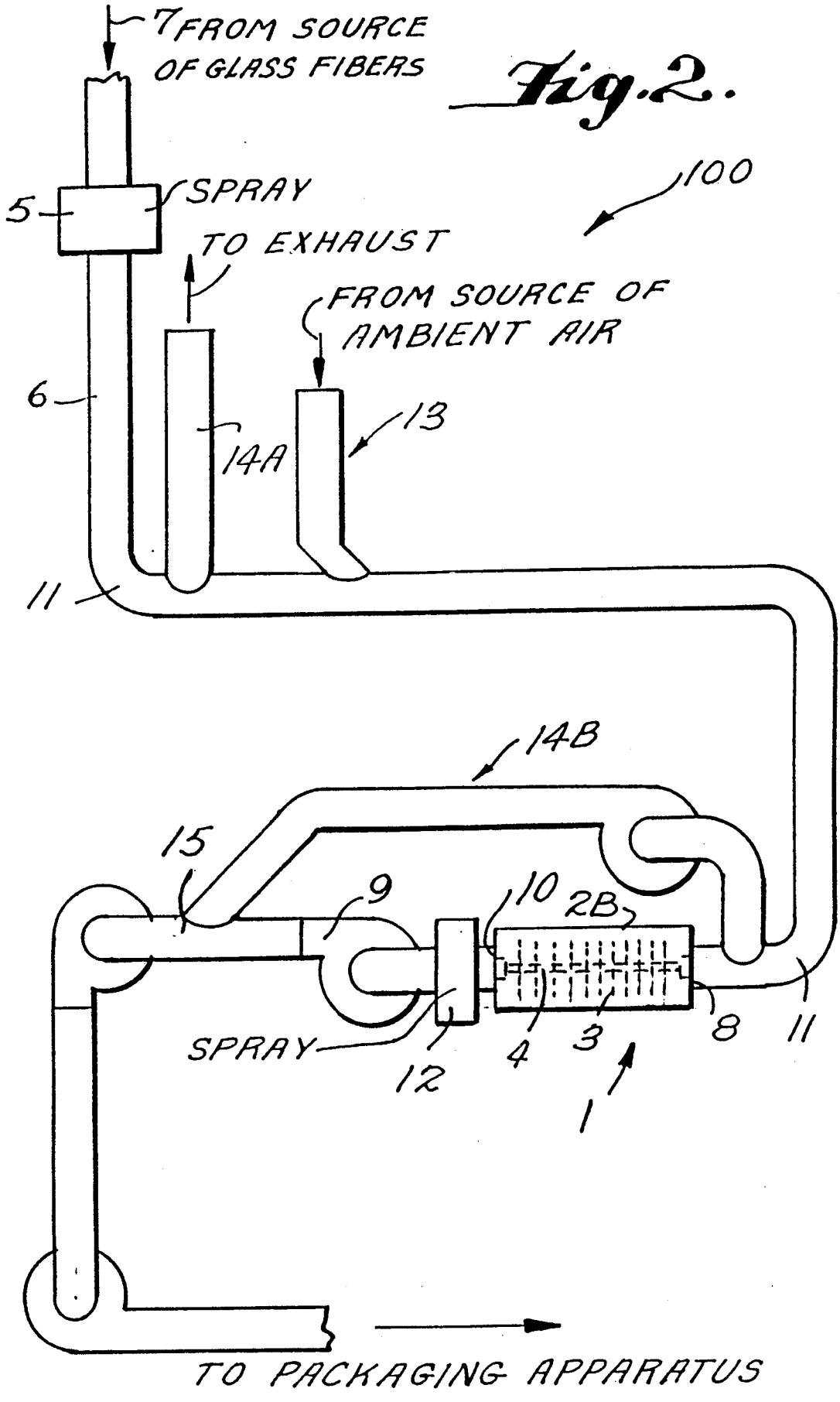


Fig. 2.


7
FROM SOURCE OF GLASS FIBERS
100
5
SPRAY
TO EXHAUST
FROM SOURCE OF AMBIENT AIR
6
14A
13
11
14B
15
9
10
2B
SPRAY
12
4
3
8
11
1
TO PACKAGING APPARATUS

LOOSE FILL INSULATION PRODUCT COMPRISING MINERAL WOOL NODULES

This is a divisional of co-pending application Ser. No. 307,495 filed on Feb. 8, 1989.

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to the manufacture of loose fill insulation products from mineral fibers.

More specifically, the invention relates to a nodulizing machine for use in an apparatus for making loose fill insulation from mineral fibers. The invention also relates to the apparatus with which the nodulizing machine is used, to a process for using the apparatus, to the use of said apparatus and to products of the use thereof.

2. Description of Related Art

Prior art loose fill insulation products and methods of making such products involve the expansion of individual pieces of feed stock, which may be, for example perlite or silicate, by means of, for example, heating. As well, the production of loose fill insulation by feeding insulation batting of mineral fibers through hammermills, Forberg mills or shredders is also known.

SUMMARY OF THE INVENTION

In contrast, the present invention teaches a loose fill insulation product and an apparatus and process for making the same, including the steps of reducing and controlling air flow, adding water to the feed stock, thereby continuously decreasing the temperature in the apparatus, and thereafter cutting and tumbling individual pieces of feed stock to produce a loose fill insulation product which comprises nodules of individual pieces, or parts thereof, of feed stock. By contact with one another in that process the individual nodules of mineral fibers are rounded.

The thermal performance of the loose fill insulation product of the present invention has been found to be greater than 18.5 RSI/m at a density of 12.8 kilograms per cubic meter. To achieve this thermal performance, the particle size may have a projected mean area of less than 12 mm^2 and a corresponding standard deviation of less than 11 mm^2 and the equivalent fiber diameter distribution or average fiber diameter is 6 microns using ASTM D 1282. The optical density of nodules of mineral fibers of the present invention are, on a linear scale ranging from 0 to 1000, less than 500 with a standard deviation of less than 150.

As is known to those skilled in the art, using a hammer mill on insulation batting of mineral fibers results in increased density of the core of each nodule, with little variation of the density of the outer portions of each nodule. As is also known to those skilled in the art, the mineral fibers of loose fill insulation products serve to retain air and it is such air that acts as the thermal insulator. Consequently, it is also known to those skilled in the art that the smaller the fiber diameters, the better the thermal performance of insulation made from such fibers. However, as those skilled in the art know, as desired fiber diameter decreases the difficulties of producing those fibers increase. Therefore, by not subjecting the mineral fiber feedstock to a device such as a hammer mill, which compresses the fibers, fibers of larger diameter may be used pursuant to the present invention to achieve thermal performance at much lower product density than is generally obtained by prior art loose fill insulation products at equivalent average fiber diameters. Moreover, the nodules of mineral fibers of the loose fill insulation product of the present invention have good uniformity, which, in turn, enhances the packing characteristics of the installed insulation.

Accordingly, the present invention provides a nodulizing machine for use in an apparatus for making loose fill insulation from mineral fibers, comprising an input portal, a conduit containing a plurality of blades and an output portal.

According to a further aspect of the present invention there is provided an apparatus for making loose fill insulation from mineral fibers, comprising a duct extending from a source of said mineral fibers to a nodulizing machine, a nodulizing machine and flow rate control means to control the flow rate of air and mineral fibers in said apparatus.

In accordance with the present invention a process for making loose-fill insulation from mineral fibers, comprises using flow rate control means to feed the fibers from a fiber source, through a duct and to a first spray apparatus, spraying a lubricity-cooling agent from the first spray apparatus onto said mineral fibers, using the flow rate control means to convey the mineral fibers subjected to the spraying to a nodulizing machine, wherein the mineral fibers are cut and tumbled to form nodules of mineral fibers by contact with a plurality of blades of the nodulizing machine, and by rubbing upon each other as they are conveyed radially, circumferentially and axially through the nodulizing machine from its input portal to its output portal by use of the flow rate control means.

In accordance with the present invention there is also provided a loose fill insulation product comprising nodules of mineral fibers.

BRIEF DESCRIPTION OF THE DRAWINGS

Embodiments of the present invention will now be described with reference to the accompanying drawings, in which:

FIG. **1** is an elevation view of an embodiment of the nodulizing machine of the present invention and

FIG. **2** is a schematic view of an embodiment of the apparatus of the present invention.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The nodulizing machine **1** of the present invention may further comprise blade motion control means **2A** Such blade motion control means **2A**, such as a variable-speed motor in V-belt drive, are capable of rotating the shaft **4** to which the blades **3** are rigidly or pivotably connected. Pivotable connection as between the shaft **4** and each blade of the plurality of blades **3** is a means, in addition to varying the speed of the motor, to control the throughput of feedstock and output of nodules. Such control is desirable in the event of a sudden increase in the input flow rate of feedstock into the nodulizing machine.

The apparatus **100** of the present invention may further comprise a first spray means **5** located between walls of the duct **6**, (which passes from the source (not shown) of the mineral fibers to the nodulizing machine **1**) between the input portal **7** of the duct **6** and the input portal **8** of the nodulizing machine **1**. The first spray means **5** is purposed to spray a lubricity-cooling agent, such as a water-silicone oil emulsion, onto the mineral fibers. (In order to facilitate the transport of nodules of

mineral fibers from the nodulizing machine **1** to a packaging apparatus (not shown), the apparatus **100** may further comprise a pneumatic conveying line **9** directed outwardly from the output portal **10** to the nodulizing machine **1**.)

In order to facilitate reduction of air flow rate in the nodulizing machine **1** at least one degassification elbow **11** of a plurality of degassification elbows may be included in the flow rate control means of the apparatus **100**. Details of the construction and use of each such degassification elbow **11** are disclosed in a co-pending U.S. patent application Ser. No. 307,520, now U.S. Pat. No. 4,971,604 filed concurrently herewith the disclosure of which is incorporated herein by reference. As well, an ambient air input means **13** may be between the exhaust conduit means **14A** of the first degassfication elbow **11** and the second degassification elbow, which may be disposed between that ambient air input means **13** and the input portal **8** of the nodulizing machine **1**. Preferably, each degassification elbow **11** and ambient air input means **13** is part of the duct **6** (extending from a source (not shown) of mineral fibers to the nodulizing machine **1**) and is located between the first spray apparatus **5** and the nodulizing machine **1**. The exhaust conduit means **14** of each degassification elbow **11** ma be connected to the pneumatic conveying line **9** at one or more re-entry portals **15** to facilitate the transfer of nodules of mineral fibers from the nodulizing machine to a packaging apparatus (not shown).

The apparatus **100** may further comprise a second spray apparatus **12** directed between the walls of the above-mentioned pneumatic conveying line **9** and capable of spraying onto the nodules of mineral fibers an agent which, when applied to the nodules of mineral fibers, is capable of retaining mineral dust. Mineral oil is an example of such a dust retention agent.

Water-silicone oil emulsion comprises an example of the lubricity-cooling agent referred to above respecting the process of the present invention. The flow rate control means used in the process of the present invention may include a plurality of degassification elbows **1** (disclosed in the abovementioned co-pending U.S. patent application) and a cool air input means Each such degassification elbow **11** and cool air input means **13** is preferably located downstream from the first spray apparatus **5** and upstream from the nodulizing machine The process of the present invention may include two further steps. The first step comprises conveying the nodules of mineral fibers from the output portal **10** of the nodulizing machine to a second spray apparatus **12**, which is capable of spraying onto the nodules of mineral fibers an agent which when applied to the nodules of mineral fibers is capable of retaining mineral dust. The second further step is the spraying of mineral dust retaining agent onto the nodules of mineral fibers. A pneumatic conveying line **9** downstream from the nodulizing machine **1** is an example of a means which may be used according to the process of the present invention for conveying the nodules of mineral fibers in the above-mentioned first further step. Mineral oil comprises an example of the above-mentioned mineral dust retaining agent.

The rates and quantities of flows of air and of fibers into the apparatus **100** and thence into the nodulizing machine **1**, together with the rate of rotation of the shaft **3**, determine the residence time of the fibers in the nodulizing machine **1** and hence determine the size, shape and thermal performance of the nodules of mineral fibers produced.

Surfaces of the nodules of mineral fibers of the loose fill insulation product of the present invention may bear an agent capable of retaining mineral dust. Mineral oil comprises an example of an agent capable of retaining mineral dust.

It will be obvious to those skilled in the art that the scope of the present invention is not restricted to the embodiments disclosed above and may be varied within the scope of the following claims without departing from the spirit and scope of the invention.

I claim:

1. A loose fill insulation product comprising mineral wool nodules and produced by a process comprising the steps of:

spraying a lubrication-cooling agent onto a flow of discrete mineral wool fibers;

pneumatically conveying the sprayed fibers at a controlled rate into a nodulizing means, said rate being controlled by conveying the sprayed fibers through at least a first degassification elbow;

cutting and tumbling the fibers into nodules in the nodulizing means by contact with a plurality of blades rotating at a controlled speed and by rubbing upon each other as the fibers travel radially, circumferentially and axially through said nodulizing means; and

pneumatically conveying the nodules from said nodulizing means,

said nodules having a thermal performance of at least 18.5 RSI/m at a density of 12.8 kg/m^3 and a particle size having a projected means area of less than 12 mm^2, with a corresponding standard deviation of less than 11 mm^2.

2. A loose fill insulation product as claimed in claim **1**, wherein the optical density of said nodules is, on a linear scale ranging from 0 to 1000, less than 500, with a corresponding standard deviation of less than 150.

3. A loose fill insulation product as claimed in claim **1**, wherein said nodules of mineral fibers have surfaces bearing an agent capable of retaining mineral dust.

4. A loose fill insulation product as claimed in claim **3**, wherein said agent comprises mineral oil.

* * * * *